The application of neural network and statistical modelling to the scale-up and optimisation of a new adhesive tape manufacturing process.

Grant-Abban, J. D

The Application of Neural Network & Statistical Modelling to the Scale-Up and Optimisation of a New Adhesive Tape Manufacturing Process

By

J.D. Grant-Abban

A thesis submitted to the University of Wales

for the degree of Master of Philosophy

2002

Department of Mechanical Engineering

University College of Swansea

University of Wales

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

This is to certify that neither this thesis, nor any part of it, has been presented, or is concurrently submitted in candidature for any degree of any university.

Candidate

Certificate of Research

This is to certify, except where specific reference to other investigations is made, the work described in this thesis is the result of the investigations of the candidate.

Candidate

.....

Director of Research

Acknowledgements

I would especially like to thank my supervisor Dr T.C. Claypole for his guidance and support during this project. At many stages in the course of this research project I benefited from his expert advice, particularly so during his careful editing which helped transform my efforts in to this thesis.

I wish to acknowledge the expertise and professionalism of the Welsh and international members of the BT1 team, through whose efforts the production facilities mentioned in this thesis were established, operated and managed. Their commitment to thorough and effective product and process understanding made this research possible. I would especially like to thank Clarke Hall Jnr. for readily sharing his neural network experience and being a patient mentoring influence whilst I was a novice getting to grips with some of the intricacies.

I am very grateful to my immediate work supervisors for their encouragement to take on this Mphil and for their sponsorship.

Finally I would like to thank several close friends for their counsel, encouragement and prayers during the tough times: most of all my best friend and wife, Jane. I thank her for her spiritual and very practical support throughout this endeavour. Without her compensating, family, career or study commitments would have slipped unacceptably.

Summary

This thesis is concerned with the successful application of statistical and neural modelling techniques to the efficient scale-up and optimisation of a new pressure-sensitive tape manufacturing facility. This thesis describes the generation of modelling data, the use of back propagation neural networks to model properties, and the optimisation of the process using the neural models.

Modelling data was purposely generated in a series of trials, using the structure of Design of Experiments to cover the process envelope systematically and efficiently. A neural model, using data from pilot-scale process experiments, was used to influence the design of the full-scale processes. Once the full-scale processes were established, a more detailed 34 factor neural model was developed. This second neural model was validated and used to optimise and explore the full-scale process. The validation exercise detected anomalous data using statistical analysis. The anomalous data was subsequently quarantined and prevented from confusing the optimisation effort.

The detailed understanding of the process came from running DoE's on the virtual process, represented by the second neural model. These DoE's provided useful contour plots and response surface visualisations of the 'black box' neural network model, giving insights that guided optimisation and indicated the potential for tape with significantly enhanced product properties.

Contents

List of Tables

List of Figures

Glossary

Chapter 1 Introduction **1**

Chapter 2 Literature Search **4**

2.1 The application of neural networks in industry 4

2.1 The use of neural networks for process optimisation 7

2.2 Closure 9

Chapter 3 Project Overview **10**

3.1 Overview of Product 12

3.1.1 Tack 13

3.1.2 Adhesion 14

3.2 Overview of Processes 14

3.2.1 Adhesive Polymer Maker 15

3.2.2 Compounding 16

3.2.3 Coating 17

3.3 Closure 19

Chapter 4 DoE and Neural Modelling Techniques **20**

4.1 Design of Experiments Modelling 20

4.1.1 The mechanics of analysing a DoE 21

4.1.2 Statistical Filters Used by DoE to Justify Decisions and Conclusions 26

4.2 Neural modelling 31

4.2.1 Background 31

4.2.2 Back propagation neural network algorithm. 34

4.2.3 Neural Network Modelling Methodology Used 37

4.3 Closure 41

Chapter 5 Modelling the Pilot-Scale System 42

5.1 Description of the First Neural Network Model Data 42

5.1.1 Application of Neural Network Modelling Methodology as per the Theory and the Exploitation of the Results 44

5.1.2 Impact of the findings on the final process design 58

5.2 Closure 58

Chapter 6 Modelling the Full-Scale Compounding and Coating Sub-Processes 60

6.1 The full-scale adhesive-polymer-metering system. 62

6.1.1 System Description 62

6.1.2 Investigation study 65

6.1.3 Analysis of the DoE : ANOVA 69

6.1.4 Implications of the experimental study 74

6.2 Full-scale resin and additive metering system 81

6.2.1 System Description 81

6.2.2 Experimental methods 86

6.2.3 Resin-metering System Characterisation 90

6.3 UV curing 96

6.3.1 UV Curing System Description 96

6.3.2 UV Radiation Modelling 98

6.3.3 Derivation of a model for specifying full-scale process equipment. 100

6.3.4 Derivation of a model involving the % lamp power 108

6.4 Closure 113

Chapter 7 Modelling the Production System 114

7.1 Description of Second Generation Neural Network Model 114

7.1.1 Application of Neural Network Modelling Methodology 116

7.2 Closure 151

Chapter 8 Discussion 153

Chapter 9 Conclusions and Recommendations for Future Work 156

9.1 Conclusions 156

9.2 Future Work 156

Bibliography

References

Appendices

{ rd "C:\\Documents and Settings\\uk022767\\My Documents\\Data\\Active\\MPhil\\Final\\Part 2.doc" }

List of Tables

Table 1 Summary of important differences between the pilot and full-scale processes 17

Table 2 Expanded DoE matrix showing interactions plus calculated effects for each factor and interaction. 23

Table 3 Data for the adjustment term calculation for factor C 27

Table 4 ANOVA table for the doe example 28

Table 5 ANOVA for the linear regression of predicted vs actual responses 49

Table 6 Taguchi L54 array. 51

Table 7 Factor names, name codes and uncoded factor levels for the L54 Taguchi DoE 52

Table 8. Analysis of Adhesion showing the effect at each level, the delta, its rank and the ANOVA. 54

Table 9. Analysis of Rolling Wheel Tack showing the effect at each level, the delta, its rank and the ANOVA. 54

Table 10 Central composite DoE factors, factor levels and responses 68

Table 11 ANOVA for the quadratic model predicting throughput 69

Table 12 ANOVA for the linear equation predicting melt temperature 72

Table 13 Results of measuring UVA dose at various belt speeds. 102

Table 14 ANOVA for the single factor model of inverse dose 103

Table 15 Lamp types available, and their configurations that allow a web 320mm wide, travelling at a line speed of 50m/min, to receive at least 0.8J/cm² . 106

Table 16 Central composite DoE and results. 109

Table 17 ANOVA for dose 109

Table 18 ANOVA for web temperature 111

Table 19 ANOVA for the linear regression of predicted vs actual combines response 121

Table 20 Factor codes, names and levels 135

Table 21 ANOVA for virtual adhesion derived form the interrogating central composite DoE run on the second generation neural network model. 136

Table 22 ANOVA for virtual liner-side rolling-wheel tack derived form the interrogating central composite DoE run on the second generation neural network 137

Table 23 DoE Equations predicting virtual adhesion 144

Table 24 DoE Equations predicting virtual rolling-wheel 145

Table 25 The sub-optimum levels of terms of the quadratic equations describing virtual adhesion and virtual rolling-wheel tack . 146

Table 26 Conditions that yield optimum properties according to DoE analysis of the neural network model 150

List of Figures

Figure 1 Screening methods decision tree 7

Figure 2 Summary of the project phases 10

Figure 3 Tape construction 12

Figure 4 Rolling wheel tack test apparatus 13

Figure 5 Map of the generic complete process 14

Figure 6 Map of the generic complete process, highlighting the polymer maker 15

Figure 7 Map of the generic complete process, highlighting compounding 16

Figure 8 Map of the generic complete process, highlighting coating 17

Figure 9 Pilot-scale coating process 18

Figure 10 Full-scale coating process 18

Figure 11 A 2 –level, 8 run full-factorial DoE with 1 response showing the levels of each factor and a response value for each of the experimental runs. 22

Figure 12 Normal probability plot for the experiment results. 25

Figure 13 Possible fits to noisy data taken from a smooth trend 33

Figure 14 Artificial neuron, a mathematical algorithm for modifying inputs in to outputs. 34

Figure 15 Back Propagation neural network architecture 35

Figure 16 Graph of the RMS error vs training cycles for differning numbers of hidden layer neurons 39

Figure 17 Summary description of data sets that fed the 1st generation neural network 43

Figure 18 plot of RMS error vs Learning cycles for different numbers of hidden layer neurons 46

Figure 19 Correlation plots of predicted and actual measurements for Adhesion and rolling wheel tack. 48

Figure 20 Ranked P values in order of decreasing significance for factors A-Y for adhesion and tack 55

Figure 21 Normal probability plot for Adhesion. 56

Figure 22 Normal probability plot for Rolling Wheel Tack. 56

Figure 23 Main effects plots for the significant Factors in the Taguchi L54 interrogation of the first generation neural network mode. 57

Figure 24 Map of the generic complete process highlighting the polymer-metering process 62

Figure 25 Schematic of the adhesive-polymer-metering system. 62

Figure 26 Adhesive-polymer-metering system heater arangements, temperature sensing points and pressure sensing and control points. 63

Figure 27 Schematic of the adhesive-polymer-metering system with some details about the factors used to generate data for the central composite DoE and the temperature and pressure measurement points. 67

Figure 28. XY graph showing the correlation between predicted and actual throughputs based on the quadratic model. 71

Figure 29 XY graph showing the correlation between predicted and actual melt temperatures based on the linear model. 73

Figure 30 Graph of predicted vs actual throughputs during model confirmation runs 75

Figure 31 A plot of all of the throughput standard deviation data collected for the DoE 76

Figure 32 Contour plot of throughput prediction generated from the model derived from the central composite DoE. 77

Figure 33 Response Surface throughput prediction generated from the model derived from the central composite DoE. 78

Figure 34 Graphical representation of the linear model for melt temperature. 80

Figure 35 Map of the generic complete process, highlighting resin addition 81

Figure 36 Schematic of the resin and additive feeding system. 82

Figure 37 Schematic of a loss-in-weight feeder showing control and feedback 83

Figure 38 Schematic showing the direction of flow of information between the SCADA and the resin-metering system. 85

Figure 39 Sample I-MR chart 87

Figure 40 The eight tests for special causes 88

Figure 41 I-MR Control charts for feeder A and "two sided test alternative" t-test for feeder A 91

Figure 42 I-MR Control charts for feeder B and "two sided test alternative" t-test for feeder B 92

Figure 43 I-MR Control charts for the additive feeder and "two-sided test alternative" t-test for the additive feeder 93

Figure 44 Correlation between feeder set-point throughputs and throughputs reported to SCADA for feeders A and,B and the additive feeder. 94

Figure 45 Top view of a curing chamber. 97

Figure 46 Schematic of the relationship between the SCADA and the full-scale UV system. 98

Figure 47 Laboratory-scale UV system from above 98

Figure 48 UV Power Puck 100

Figure 49 Schematic of the laboratory UV curing system. 101

Figure 50 Schematic showing the laboratory UV curing system and the full-scale UV curing system. 102

Figure 51 Relationship between line speed and reciprocal dose using the laboratory-scale 120W/cm system. 104

Figure 52 Plot of predictions of different lamp types in different configurations 107

Figure 53. Correlation plot of the model prediction and the actual values measured at various line speeds with different configurations of lamps. 108

Figure 54 Contour plot for dose prediction generated from the model derived from the central composite DoE. 110

Figure 55 Contour plot for web temperature prediction generated from the model derived from the central composite DoE. 112

Figure 56 Contour plot of the model showing the process conditions required to achieve the target dose whilst not producing excess UV. 113

Figure 57 Summary description of data sets that fed the second generation neural network 115

Figure 58 plot of RMS error vs Learning cycles for different numbers of hidden layer neurons 119

Figure 59 Correlation plots of predicted and actual measurements for the combined responses in the training data set. 120

Figure 60 Normal probability plot of the residuals of the neural network model predicting the test data set. 125

Figure 61 Dotplot of the residuals of the trained neural network model predicting the test dataset for each subset of the combined response. 126

Figure 62 Normal probability plot of the residuals of the neural network model predicting the test data set with the face-side rolling wheel data and outlier data removed. 127

Figure 63 Histogram of the residuals of the test data set with face-side rolling-wheel and outlier data removed, with the 95% t-confidence interval of the mean. 128

Figure 64 95% Confidence intervals of the standard deviations for the subsets of the combined response. 129

Figure 65 Plot of the residuals verses combined response level. 130

Figure 66 XY Plot of the virtual responses (liner-side only) for the central composite DoE data 131

Figure 67 XY plot showing the degree of correlation between virtual adhesion and reciprocal virtual rolling-wheel data generated for the central composite DoE 132

Figure 68 ANOVA of linear regression of inverse tack and adhesion 132

Figure 69 XY plot of the residuals from the linear equation verses the prediction of virtual reciprocal rolling-wheel. 133

Figure 70 Virtual adhesion main effects plot for each factor in the central composite DoE, together with a key. 139

Figure 71 Virtual rolling-wheel tack main effects plot for each factor in the central composite DoE, together with a key. 140

Figure 72 XY Plot of the virtual responses (liner-side only) for the central composite DoE data with target requirements superimposed. 143

Figure 73 Histogram plot of the first 56 production test results for liner side adhesion. 147

Figure 74 Histogram plot of the first 56 production test results for liner side rolling wheel. 148

Figure 75 Contour plot of virtual liner-side adhesion showing predictions at the extremes of polymer age at the time of coating, tested fresh off the machine. 149

Figure 76Contour plot of virtual liner-side rolling-wheel tack showing predictions at the extremes of polymer age at the time of coating, tested fresh off the machine. 149

Figure 77 XY Plot of the virtual responses (face-side rolling-wheel data removed) for the central composite DoE data with the 3 super property predictions 151

Glossary

Additive	=	A constituent of the compounded adhesive that causes polymer cross-linking when energised with UV
Adhesion	=	The force of attraction between two materials
Adhesive polymer maker	=	A process that polymerises monomers in to a major raw material for the coating and compounding processes
Coating	=	A process for the application of compounded adhesive on to a web/ substrate which is wound in to jumbo reels
Compounding	=	A process for the blending of chemical constituents using metering feeders and a twin screw extruder in to a homogeneous melt of adhesive
Curing	=	Chemical cross-linking of adhesive polymer with the aid of additive and UV radiation
Die (coating)	=	A unit that provides a shaped orifice, out of which the melt is forced under pressure.
DoE	=	Design of Experiments, a statistically based multi-variable modelling technique
Hotmelt	=	A thermoplastic adhesive system coated without solvents, at elevated temperatures.
I-MR	=	Individuals and Moving Range statistical process control chart
Neural modelling	=	A modelling technique based on artificial neural network algorithms
Neural network (artificial)	=	A data processing architecture based around models of biological systems that requires no prior knowledge of the system being modelled
Resin	=	A material that modifies adhesive polymer Tack and adhesion properties
SCADA	=	Supervisory Control And Data Acquisition; a PC-based man-machine interface and data logger

		man-machine interface and data logger
Sub-process	=	A process can be broken down in to its constituent sub-units that themselves could be considered processes in their own right, i.e. they convert inputs in to outputs. These are referred to as sub-processes
Tack	=	The ability to bond to substrates with very little applied pressure.
UV	=	Ultra-violet (radiation)

Chapter 1 Introduction

Any new product introduced into the market has to be manufactured with a consistency and quality of performance worthy of the brand name. Increasingly this has to be achieved against a commercial background that forces the product to be introduced in a shorter time scale. Traditional routes of process development with long trials to develop experience of the product prior to launch are not acceptable. It may be possible to engineer for quality by making assumptions based on the experience of other established products or of mature processes, but where the products are unique, incorporate new technology (that may even be proprietary) and require new processes to manufacture them, there is by definition little or no useful operational experience on which to base an engineering design. It is only through adequate early understanding of the likely product and process interactions that data-based decisions can be made about the final engineering process designs. Understanding the necessary level of control needed to ensure robust quality without incurring the unwelcome capital expense and potential time over-runs of an over-engineered machine may mean the difference between success and failure.

This thesis is concerned with the application of statistical and neural modelling techniques to establish a totally new manufacturing facility for a coated pressure sensitive tape with aggressive project timing and cost constraints. An essential element for achieving this was the efficient use of data from pilot-scale trials to model and design the full-scale processes and the development of a "virtual process" on which to experiment to gain insights into what the likely optimised process conditions would be and how to improve product performance while keeping costs down by generating minimal amounts "real process" trial data. All of these were to ensure that the level of risk at product launch was acceptable to the organisation.

The system studied in this thesis has modelled 34 factors in 200 experimental runs. Traditional statistical modelling methods such as Design of Experiments, DoE, would have been incapable of modelling all these variables to the level of detail attained without having to run an impractically large number of experimental runs. The alternative to large experiments when modelling large numbers of variables using DoE methods is to lose model detail by forcing a number of the variables to be 'frozen' at constant levels, resulting in lack knowledge about the effects of altering those variables. The resulting model would therefore offer no insight in to the potential benefits of being able to adjust those variables during optimisation.

The approach taken in this thesis means that the resulting verified model is neural rather than statistical. The understanding of the process comes from interrogating the neural model with DoE's. DoE's are run on the virtual process, represented by the neural network model. Data visualisation tools in common use with DoE, such as contour plots and response surfaces, convert the 'black box' neural network in to informative representations of the system. Keeping the model neural rather than statistical has the advantage that the model still takes in to account the influences and interactions of some of the factors with minor influence, necessarily frozen at a constant level in the statistical approach. Another argument for leaving the model in neural form is that if any of these minor frozen variables change for some reason, it does not necessitate a new DoE to be run to get a new valid statistical model. The neural model is robust to imposed and unforeseen changes in minor factors.

The published literature on the application of neural networks and statistical methods to the optimisation of this type of production process is reviewed in chapter 2. The phases of the project from pilot-scale to full-scale production, together with the tape is described in chapter 3. The method of developing neural networks for this type of problem solving is described in chapter 4. The application of neural networks to this particular process is described in three chapters. Chapter 5 describes the modelling of

the pilot facilities, chapter 6 looks at the scaling to full-scale production and chapter 7 describes the model of the complete system. The overall findings are discussed before conclusions and recommendations for future work are made.

Chapter 2 Literature Search

In this review, two aspects of the literature are explored. The first is a summary of the status of neural networks as an industrial tool. The second is a comparison of the neural methods used in the research with alternative methods. It should be noted that this project was completed in 1997 and should be considered against a background of the available literature at the time. However this work is set in context in this review by reference to work that has subsequently been published. The use of neural networks as an industrial tool is a relatively young subject, and little of any direct relevance could be found whose publish date preceded the start of the research. Where researchers have used neural methods where traditionally statistical methods with their limitations would have been used, they have done so with a fundamentally different philosophy.

2.1 The application of neural networks in industry

J.G. Taylor's book[1], published in 1993, has a title reflecting the then existing unrealised potential of neural networks to impact areas of industry. His book, *The promise of Neural Networks,* summarises the level of deployment of neural networks as practical solutions to industrial opportunities. He lists eight areas in which neural networks had been applied, together with some specific examples of their deployment; although none of these are similar to the application reported in this thesis. The eight areas are as follows:-

a) **Pattern recognition and classification** e.g. handwriting recognition, sequencing of amino acids

b) **Speech recognition**

c) **Picture recognition**, identification of objects on a conveyor, quality control, automatic satellite image analysis

d) **Robotics**, e.g. sensor-actuator coordination, stabilising problems.

e) **Signal processing,** e.g. reconstruction of degraded telephone signals

f) **Optimisation**, e.g. optimisation of waiting times of flight crews between connecting flights.

g) **Control**, e.g. catalyst addition in a chemical plant

h) **Time series prediction**, i.e. prediction of future values e.g. on the stock market

The area closest to matching the project in this thesis is area 'f', optimisation. No manufacturing optimisation examples were given in the list. This implies that industrial process modelling for optimisation purposes, was either unknown or too much of a fringe application to mention. Taylor recognises the potential relevance to industry of this optimisation area to grow. Taylor goes on to make several predictions about the shorter-term future that include a reference to deployment in a factory. He predicted that neural networks' futures lay in games, domestic appliances, factory control for a few parameters and few-variable time series predictions. He considered the major benefit of neural networks to factories lay in controlling a few processes parameters in real time, longer-term predictions of growth areas for deployment included automatic vision and speech recognition. Neither his shorter-term nor his longer-term prophecies included using neural networks as a modelling technique to make scale-up and optimisation of complicated processes an efficient exercise. Taylor's analysis of where neural networks were already deployed and their likely growth areas, are consistent with other literature from that decade. Bishop[2] agrees that pattern recognition is the dominant category of application. He considered that neural network applications, where successful, have benefited from a mature and principled approach, instead of the ad hoc approach of some earlier applications. His authoritative textbook demonstrates the level of maturity of the field of neural networks by giving them a solid statistical foundation. Lionel Tarassenko[3] comments that the principled approach that Bishop describes has made possible many solutions to pattern recognition or signal analysis problems. He lists similar areas of neural network deployment to Taylor but also recognises the existence of neural computing in process modelling in manufacturing. He does so in a tabular summarisation of

typical business functions. In 1998, Fogelman and Gallinari[4] commented that neural network technology was understood, software was available and some successful case histories existed, yet neural networks were not in use in every area of industry or the service sector. They presented 10 sections, similar to those identified by Taylor, covering 10 areas of neural network activity in industry. The category with the closest fit to the work reported in this thesis is the section entitled 'Process Engineering, Control and Monitoring'. This deals with using neural networks in real time to react to dynamic process conditions. However they did not report any examples that approximate to what was done in this project with neural networks, i.e. using the neural network model, off-line, to direct engineering solutions to scale-up process issues. Raj et al[5] worked on an extrusion process that, with the aid of neural network modelling, enabled them to dramatically reduce the level of human intervention and experimentation required to achieve quality product when starting to make initial lots of new parts. Their neural model allowed them to predict likely best process conditions for new parts using data gathered from previous parts. Their work on an extrusion process, although producing a fundamentally different product to the one researched in this thesis, demonstrates that the marriage of real process data with neural network modelling technology has lead to practical manufacturing benefits rather than just academically interesting results. Moser et al[6] also demonstrated that neural network modelling of product quality using process input data lead to optimised product performance. Their work was carried out on plasma polymerised thin film functional quality made on a web-based process; in this respect their process is similar to the process researched in this thesis. Chen et al[7] carried out work on a resin polymerisation process; the resin is similar to one of the raw materials used in the research for this thesis. In their neural network deployment, they describe the advantages of a neural network prediction over non-linear regression when used to control the melt flow index of a resin. The advantages lie in the ability of the neural network to adapt to process drifts.

2.1 The use of neural networks for process optimisation

Trocine and Malone compare screening designs for finding relationship between factors and responses[8]. They list several statistical techniques, their advantages and disadvantages and come up with a decision tree for which technique to use under general circumstances (Figure 1).

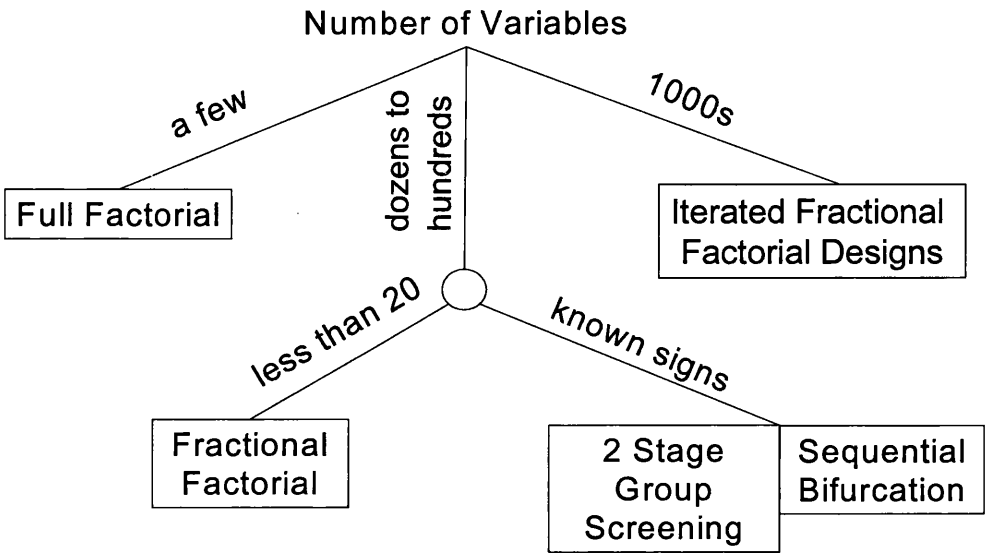


Figure 1 Screening methods decision tree

This decision tree, if applied to the screening problem posed in this thesis, would have resulted in a less than satisfactory solution. The number of variables eventually experimented with was 34 with unknown signs. 'Known signs' are defined by Trocine and Malone as existing knowledge about the effects of factors on a response that allow groupings to be made of variables that alter the response in a similar direction i.e. going from low to high levels of each factor in the group alters the response in the same direction. The decision tree does not really fit an option of >20 variables of unknown sign. The use of neural networks to estimate the effects of >20 variables, was not considered by journal. Normal statistical methods have not been applied in

situations where the effects of more than 20 variables are unknown. Vaidyanathan et al[9] compared response surface methodology with neural networks when applied to the modelling and optimisation of rocket engine parts. They compared Response Surface Methodology, rsm, with two different types of neural network architecture for prediction accuracies including back propagation. However they only considered a limited number of factors, 3, with limited data points, 45. The rsm and neural network modelling produced comparable results. In 2001 Nandi et al[10] report achieving a successful optimisation of a piece of chemical plant using a neural network model derived from input-output data from that plant. They used genetic algorithms and simultaneous perturbation stochastic approximation to interrogate the neural network, leading them to their optimisation goal. They claimed that their approach is applicable to all kinds of process design and optimisation problems.

Thomas Pyzdek[11] collated pre-existing data held in manufacturing IT databases, modelled the appropriate data with a neural network, validated it with data not used in neural network training, ran DoE's on the virtual process to find optimum settings, ran a DoE on the real process at those settings suggested by the optimised virtual process. Because the results concurred, he was able to move along the path towards real-process optimum settings. The target of Pyzdek's approach was to develop a verified statistical model. This statistical model is the result of a real DoE run on a real process. The real process is therefore understood in the context of an equation that includes terms representing a few factors with the rest of the process variables held 'frozen' at a constant level. If the number of influential variables is large then statistical methods involving DoE would necessarily have to select a limited number of those variables to become experiment factors. This is because the number of experimental runs required by DoE methodology increases exponentially with the number of factors, a real constraint for an organisation with finite resources. This method is not robust to imposed and unforeseen changes in the frozen variables; a new DoE would have to be run to get a new valid statistical model. Chu et al[12]

proposed an optimisation methodology using neural networks in 2002. They demonstrated their approach by optimising a simulation of a coal-fired boiler. They ran an iterative cycle of structured experiments, based on orthogonal arrays, on the simulator and combined all the resulting data in a neural network. They then used the neural network to predict optimum properties. They tested the validity of the prediction by running the proposed optimum settings on the boiler simulator. If the results were unsatisfactory, the cycle was repeated until a satisfactory optimisation was achieved. Some of their approach is similar to the one taken in the research for this thesis. Their use of orthogonal arrays to generate data for neural network training uses the same philosophy as that adopted in this thesis, i.e. needing to efficiently cover design space giving maximum insight for minimum expenditure of resource.

2.2 Closure

The problem described in this thesis involves the simultaneous optimisation of a large number of variables (34). This makes it impractical for the application of statistical DoE techniques. When the work was undertaken, the application of neural networks to manufacturing was completely novel and there has been little subsequent published work of direct relevance to this project. The next chapter describes the phases of the project, the tape and the production process.

Chapter 3 Project Overview

The project had three distinct phases as summarised below.

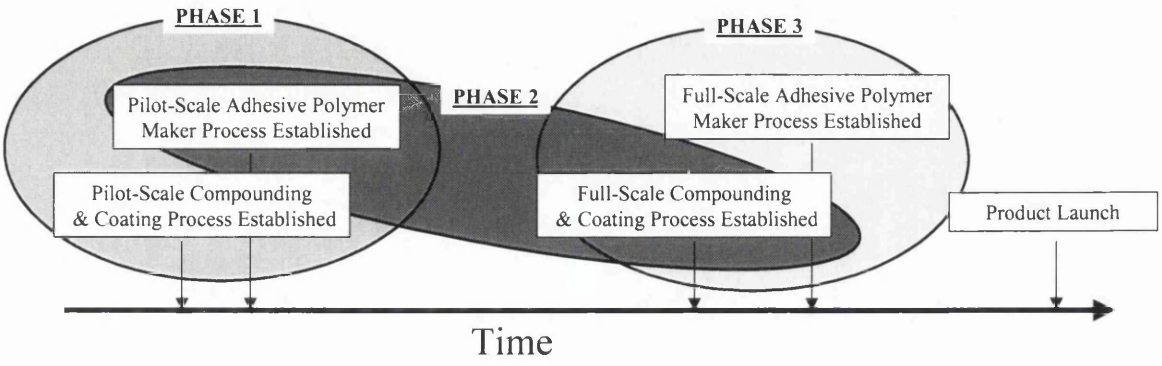


Figure 2 Summary of the project phases

The first phase was the development of the product in the laboratory using a pilot-scale adhesive polymer maker and a pilot-scale compounding and coating line. The second phase was scaling up the product from the pilot-scale compounding and coating line to the production-scale compounding and coating line, still using the raw adhesive polymer from the pilot-scale adhesive polymer maker. The third phase was full-scale production on the production-scale compounding and coating line using adhesive from the production-scale adhesive polymer maker. These three phases preceded product launch. The product launch date was fixed before any full-scale equipment existed. To meet this target a method was needed to minimise the amount of time required to learn how to make the product on full-scale equipment and to speed up qualification of the processes for product launch.

The product has a number of measurable features that describe functional performance. The magnitude of these features can usually be influenced by the raw material properties and by the manufacturing process.

The customers' perception of functional quality can be translated into tolerances or specifications for product measurables. Aligning the product with these customer derived targets and not deviating should maximise customer perceived functional quality. Product deviating from target values causes customer dissatisfaction. The greater the deviation, the greater the dissatisfaction. The ultimate goal is to optimise functional quality with process variables that are easy to set and maintain. The influence of the raw materials and manufacturing process therefore needs to be understood.

Models derived from information or data, transform the information into understanding. There are costs to deriving models because they require real data. Time and materials are consumed in generating data, important considerations for this project with limited raw materials and a tight timeline. The most detailed models can be developed by actually making and testing every possible combination of raw material variable and manufacturing process variable. However this is not the most cost effective. The alternative is to make and evaluate strategic combinations of raw material and process variables and rely on extrapolation/ interpolation between them to fill in the gaps. A good model will explain the overwhelming majority of a product response in terms of the process variables and raw material variables together with the influence of factors that are either too difficult or too expensive to vary such as ambient humidity. A good model will have adequate descriptive and predictive properties that allow proper alignment of the product to customer requirements, without exceeding the cost and time constraints.

For this reason statistical methods and neural network models were used because of their ability to find relationships in complex, non-linear, highly interacted data, as was highlighted by the literature review, chapter 2. A preliminary neural network model was used to learn from the pilot-scale processes to influence the design of the full-scale processes. Statistical methods were used to optimise sub-processes on full-scale

compounding equipment and a further enhanced neural network model was used to optimise the full-scale processes. These models, once validated, were exploited as "virtual" processes enabling the simulation of the manufacturing of new formulations or any other changes that may be required and allow the study of the possible implications. Screening of trials could take place without having to necessarily commit to real manufacturing.

This thesis first discusses analysis of the data from phase 1 pilot-scale process using a neural network model (chapter 5). This analysis was then used to shape major aspects of the design of the full-scale compounding and coating process. This required the running of Phase 2 and phase 3 experiments and analysis of the data using statistical and neural techniques throughout the project (described in chapter 7).

3.1 Overview of Product

The product is a double-sided hotmelt adhesive tape whose construction is represented in Figure 3. It consists of coatings of adhesive either side of a carrier. The adhesive is a blend of cross-linked polymer and resins giving the product its performance properties of tack and adhesion. The tape is packaged on a removable liner (which is discarded during application) and sold in various roll formats.

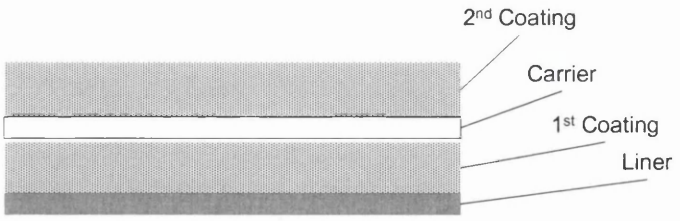


Figure 3 Tape construction

This product is sold into a variety of industries as a general-purpose high performance double coated tape. One customer bonds this tape to profile extrusions as they are being extruded and leaves the liner in place. The end-users of the extrusions remove the liner and stick them to a variety of surfaces without having to use mechanical

fastenings. The advantages to the end user are because visible mechanical fastening such as rivets, screws etc are not required, application time is reduced and cosmetics are enhanced. Another application includes using the tape to bond sailcloth together before stitching. There seems to be no typical application apart for the requirement for, high tack and high adhesion performance.

3.1.1 Tack

The tape is able to bond to substrates with very little applied pressure. This characteristic can be evaluated using the rolling wheel tack test. The test involves rolling a delrin wheel down an inclined plexiglass track (Figure 4). As it leaves the incline, it is free to carry on rolling along the horizontal base. The length of tape being tested is attached to the base. The wheel (cleaned with heptane) is released and as it accelerates down the incline builds up momentum. As the wheel leaves the ramp, it contacts the tape, which eventually stops the wheel. The tack is quoted as the distance the wheel travels along the tape. The higher the tack, the shorter the distances travelled.

Figure 4 Rolling wheel tack test apparatus

3.1.2 Adhesion

The other quality characteristic is the adhesion. It is not possible to measure adhesion directly as adhesion is the force of attraction between two materials. The only way to get a measure of the adhesion strength is to mechanically break the bond. Adhesion in this study is defined as the average force, measured by an Instron, required to peel a 25mm tape sample off a chemically clean steel plate. The sample, having previously been acclimatised to the test lab's controlled temperature and humidity, is prepared by sticking the appropriate side to the steel plate by applying a 2kg rub down force with a rubber covered roller travelling at 300mm/min, in both directions. The sample is then left for 15min to establish the bond. The steel plate is placed into the lower jaw of the Instron and clamped into position. The free end of the sample is placed into the top jaws, aligned and clamped into position ensuring that the sample is taut. The Instron then pulls the jaws apart at 300mm/min, peeling the tape off at 180°. The average force experienced by the Instron's loadcell is recorded as the adhesion test result.

3.2 Overview of Processes

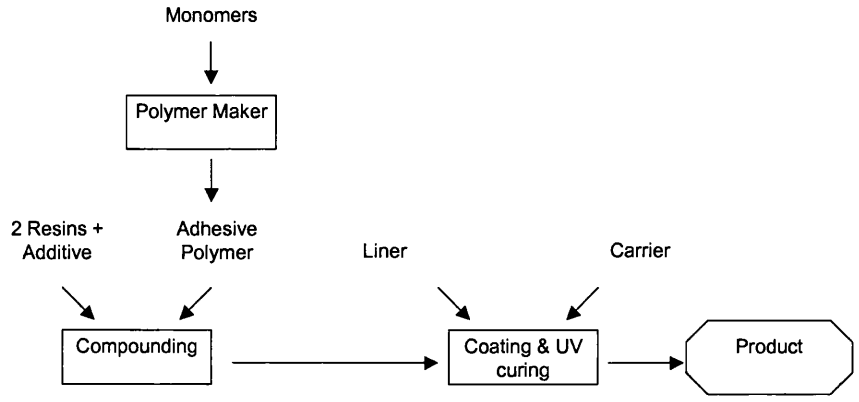


Figure 5 Map of the generic complete process

The generic process, detailed in the diagram above, combines adhesive polymer with 2 resins and an additive in a compounding process. This compounded material is continuously extruded on to web through a die. The extrudate is irradiated with UV to initiate chemical cross-linking.

3.2.1 Adhesive Polymer Maker

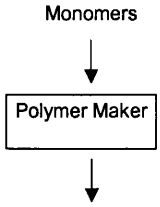


Figure 6 Map of the generic complete process, highlighting the polymer maker

The Adhesive Polymer Maker polymerises liquid monomers into an adhesive polymer gel that subsequently becomes an input raw material to the compounding process. Both pilot and full-scale processes are sited at facilities remote from the compounding and coating processes and each other. Therefore there is batch consumption of this raw material. However unlike the other input raw materials, the manufacturing process of the adhesive polymer gel is within the control of the process plant and thus forms part of this project.

The process data available on the Adhesive Polymer Maker includes process running conditions and finished goods quality measurements such as molecular weight average and distribution. This data is used by the polymerising process supervisors to control consistency. Polymer maker process conditions affect polymer properties which in turn affect tape properties. Until this investigation the exact nature of the relationship was unknown. The polymer data in isolation currently cannot predict the final tape properties. The polymer properties can be combined with coating and compounding data to predict physical properties of the tape.

The pilot line differs from the full-scale line in size. The pilot line was used for the formulation and screening type experiments as its productivities are in line with those required at the development stage.

3.2.2 Compounding

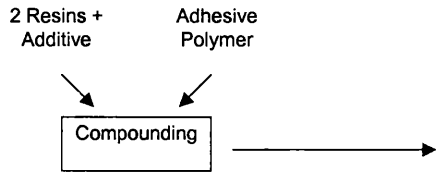


Figure 7 Map of the generic complete process, highlighting compounding

Adhesive polymer, resins and additive are fed into a twin screw compounding extruder, blended under hot conditions and extruded through a die at a coating station. The important differences between the pilot-scale process and full-scale process are summarised Table 1.

Process element	Pilot-scale	Full-scale	Implications of differences
Adhesive polymer addition	Simple volumetric system capable of 4% of full-scale system	Complex volumetric system. Delivery rate depends on several interacting factors	Full-scale system needs modelling to ensure predictable and stable delivery rate
Resin and additive addition	The resins and additive are batch blended and melted before feeding into the extruder as a molten liquid	The resins and additive are metered into the extruder via loss-in-weight feeders and melted in the extruder	Pilot-scale system has heat degradation issues not present in full-scale system.
Compounding twin screw extruder	2.5% throughput of full-scale extruder due to size.	Slightly different screw design than the pilot-scale extruder with similar rpm	Screw work effects need to be correlated
Coating die	One coating die, therefore two passes necessary to produce double coated tape	Two coating dies coating both sides of the web in one pass	Ageing effects and trapped fugitives could show as a difference between the two processes

Table 1 Summary of important differences between the pilot and full-scale processes

3.2.3 Coating

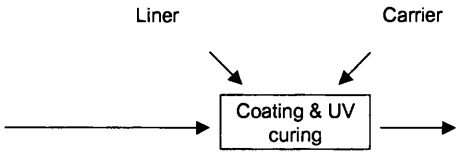


Figure 8 Map of the generic complete process, highlighting coating

The coating process is mechanically and electrically linked to the compounding process, sharing feedback, control systems and operators. Liner is unwound from stockrolls and receives a coating of compounded adhesive from the die. The coated adhesive is irradiated with UV to initiate chemical cross-linking in a curing chamber immediately after the coating station. Carrier is then laminated on to the coating. The pilot-scale coating line then winds this construction into a reel. The reel is then passed through the same process again but missing out the lamination stage, adding the second UV-cured adhesive layer (Figure 9).

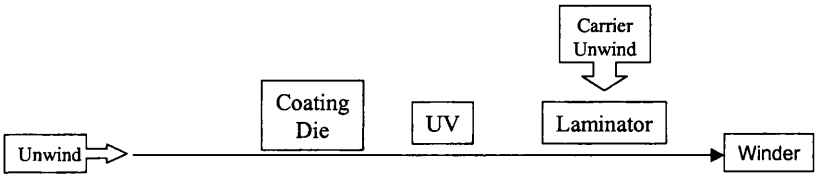


Figure 9 Pilot-scale coating process

The full-scale coating line (summarised below) has an additional coating station and an additional curing chamber before the winder eliminating the need for a second pass.

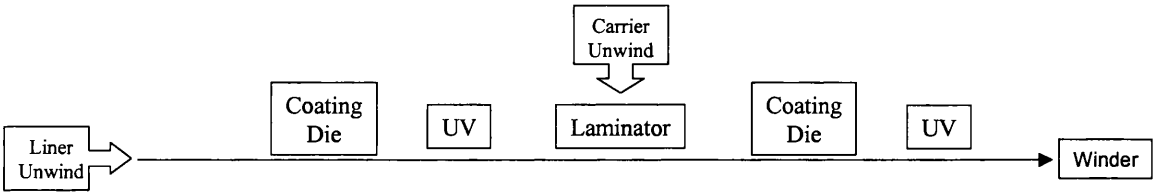


Figure 10 Full-scale coating process

Other important differences between the lines include a more than 400% increased maximum UV power output capacity for the full-scale line to cope with a 1,000% greater m^2 per hour coating capability of the whole system.

3.3 Closure

The process required to manufacture a high performance adhesive tape from an input of polymer has been described. The differences between pilot and full-scale production systems have also been highlighted. The mode of application of neural networks to the project is described in the next chapter.

Chapter 4 DoE and Neural Modelling Techniques

4.1 Design of Experiments Modelling

Design of Experiments (DoE) is a methodology that delivers efficient experimentation. It aims to maximise the amount of useful information, suitable for understanding a system, whilst minimising the number of experimental runs. It does this by ultimately deriving a mathematical model, based on factors altered during the experiment. The model is derived from running a prescribed series statistically optimised experimental runs, using combinations of factors at different factor levels (usually a high level, a low level and an average level).

DoE involves a number of stages. It starts with a basic appreciation of the system being investigated (i.e. some basic knowledge about operating the system, but not necessarily know how to get a desired outcome). Variables are selected to become factors for the experiment and factor levels assigned. The assigned high and low levels of factors need careful consideration. Inappropriate choices may lead to missed valuable information. For instance, if the factor levels are too close together, the system that gives the response may be "noisy" enough to mask the effect of changing the factor over such a relatively small amount. Too large a difference and detail between the levels will go unobserved.

Having chosen which factors are relevant, it is necessary to match the best prescribed design to the objectives of the investigation. There are a number of possible designs to choose from. Low resolution designs (small number of runs compared to the number of factors) yield potentially ambiguous models but are less costly to run. Examples are Taguchi[13] designs and Placket and Burman designs[14]. A type of Taguchi design is used in Chapter 5 where a more detailed explanation of the merits of this type of

design are given, and the circumstances that made it the design of choice. High resolution designs (higher number of runs compared to the number of factors) yield more precise models but at a greater cost. This effect rises exponentially with the number of factors. An example of a higher resolution design used in this thesis is a Central Composite DoE design. (See Chapter 6 for a more detailed explanation of why this type of design was the most appropriate choice). These designs use five levels of each factor to generate the data for modelling. The data is suitable for detailed analysis that can lead to the modelling of complex relationships between factors and factor levels. Response surfaces, 3D plots of 2 factors and a response, can be used to visualise the results (see Chapter 6, Figure 33 for an example). These designs can also predict local maxima or minima in the responses within the scope of the experiment, allowing precise optimisation.

The data from the experiment is passed through various statistical filters that determine the statistical appropriateness of including or excluding factors or combinations of factor from the final mathematical model. The filters also assess the appropriate choice of model (linear, quadratic or cubic polynomials) and give some idea of the error that is expected from using the model to predict the system's response. Throughout the analysis, various indices are taken in to account that lead to the most appropriate interpretation of the data. The resulting statistically significant models can be used to predict optimum settings for best quality.

4.1.1 The mechanics of analysing a DoE

Subsequent chapters describe the use of and results from DoE's in some detail. To avoid lengthy explanation of how the results were derived each time, a generic DoE is analysed below to demonstrate the mechanics of the analysis. This example explains how a model is generated from experimental data and how the model's integrity is assessed using statistical filters to justify conclusions.

In this example a 2 level factorial DoE was run varying 3 factors, A B & C, between low and high levels according to the prescribed matrix in Figure 11. The high and low levels, represented by +'s and –'s, were appropriately set at values likely to yield valuable information with the aid of some basic process knowledge. A response, Y, was measured for each run.

Factor	Name	Low Actual	High Actual	Low Coded	High Coded	Design Units (High - Low)
A	Temperature	90°C	120°C	-	+	30°C
B	Pressure	15bar	25bar	-	+	10bar
C	Dwell Time	10s	100s	-	+	90s

Run	Factor 1 A	Factor 2 B	Factor 3 C	Response Y
1	-	-	-	93.5
2	+	-	-	99.4
3	-	+	-	92.4
4	+	+	-	98.3
5	-	-	+	95.8
6	+	-	+	101.2
7	-	+	+	94.7
8	+	+	+	100.8

Figure 11 A 2 –level, 8 run full-factorial DoE with 1 response showing the levels of each factor and a response value for each of the experimental runs.

This matrix can be expanded to show all of the analysable interactions (combinations of factors). These are shown in Table 2. It is now possible to contrast the values of Y when the factors and interactions were at their low levels and high levels (red vs. green). If these are averaged, then the **"effect"** of going from low to high levels can be calculated for each factor and combination of factors. For example, factor C was at the low level for the first four runs and at the high level for the last four runs. Consequently when factor C was at the low level, the response values were 93.5, 99.4, 92.4, 98.3; averaging 95.9. When factor C was at the high level, the response values were 95.8, 101.2, 94.7, 100.8; averaging 98.13. The net difference in averages is 2.2. Therefore the effect of changing levels from high to low (or vice versa) for factor C is 2.22 units and is properly called the **main effect** of C.

	Main Effects			Interactions				Response	Response sorted by level						
Run	A	B	C	AB	AC	BC	ABC	Y	A	B	C	AB	AC	BC	ABC
1	−	−	−	+	+	+	−	**93.5**	93.5	93.5	93.5	93.5	93.5	93.5	93.5
2	+	−	−	−	−	+	+	**99.4**	99.4	99.4	99.4	99.4	99.4	99.4	99.4
3	−	+	−	−	+	−	+	**92.4**	92.4	92.4	92.4	92.4	92.4	92.4	92.4
4	+	+	−	+	−	−	−	**98.3**	98.3	98.3	98.3	98.3	98.3	98.3	98.3
5	−	−	+	+	−	−	+	**95.8**	95.8	95.8	95.8	95.8	95.8	95.8	95.8
6	+	−	+	−	+	−	−	**101.2**	101.2	101.2	101.2	101.2	101.2	101.2	101.2
7	−	+	+	−	−	+	−	**94.7**	94.7	94.7	94.7	94.7	94.7	94.7	94.7
8	+	+	+	+	+	+	+	**100.8**	100.8	100.8	100.8	100.8	100.8	100.8	100.8
								Mean of "-" level	94.1	97.48	95.9	96.93	97.05	96.93	96.93
								Mean of "+" level	99.93	96.55	98.13	97.1	96.98	97.1	97.1
								Effect (Δ Means)	**5.83**	**-0.92**	**2.22**	**0.18**	**-0.07**	**0.18**	**0.18**

Table 2 Expanded DoE matrix showing interactions plus calculated effects for each factor and interaction.

Having calculated the effects for each factor and combinations of factors, a method is needed for sorting out whether the effect is real, i.e. due to changing levels within the experiment, or is caused by experimental noise. Experimental noise is randomness in the response generated by inaccuracies in measurements, or settings, plus any other sources of normal variation in the response that are typical to the system. A simple yet powerful graphical technique exists to aid decision. It is a form of plot called a normal probability plot. This technique assumes that the experimental noise is random and normally distributed (normal in the statistical sense) and therefore correlates with the normal probability function. Any effects stand out from the background noise and therefore do not correlate with the normal probability function. These plots are designed to help highlight significant effects. The data for the plots is constructed as follows:-

1. List the effects of each term, rank them and place them in order. For the example in Table 2:-

Term	Effect	Rank	Order *i*
B	-0.925	1	1
AC	-0.075	2	2
AB	0.175	3	3
BC	0.175	3	4
ABC	0.175	3	5
C	2.225	6	6
A	5.825	7	7

2. Calculate the quantiles for a normal probability distribution (Q) for each ordered term (i) using the following equation: [14]

$$Q(i) = 4.91\left\{\left[\frac{i-0.375}{N+.25}\right]^{0.14} - \left\{1-\left[\frac{i-0.375}{N+.25}\right]\right\}^{0.14}\right\}$$

Where N is the total number of terms.

Carrying on with the example in Table 2 :-

	X Axis			Y Axis
Term	Effect	Rank	Order *i*	Quantiles for a Normal Distribution Function
B	-0.925	1	1	-1.365
AC	-0.075	2	2	-0.756
AB	0.175	3	3	-0.351
BC	0.175	3	4	0.000
ABC	0.175	3	5	0.351
C	2.225	6	6	0.756
A	5.825	7	7	1.365

3. Plot the effects on the X-axis and the corresponding quantiles on the Y-axis. Using estimation, align the terms most likely to be due to noise. Points significantly off that line indicate the calculated effect is likely to be real. For the example in Table 2:-

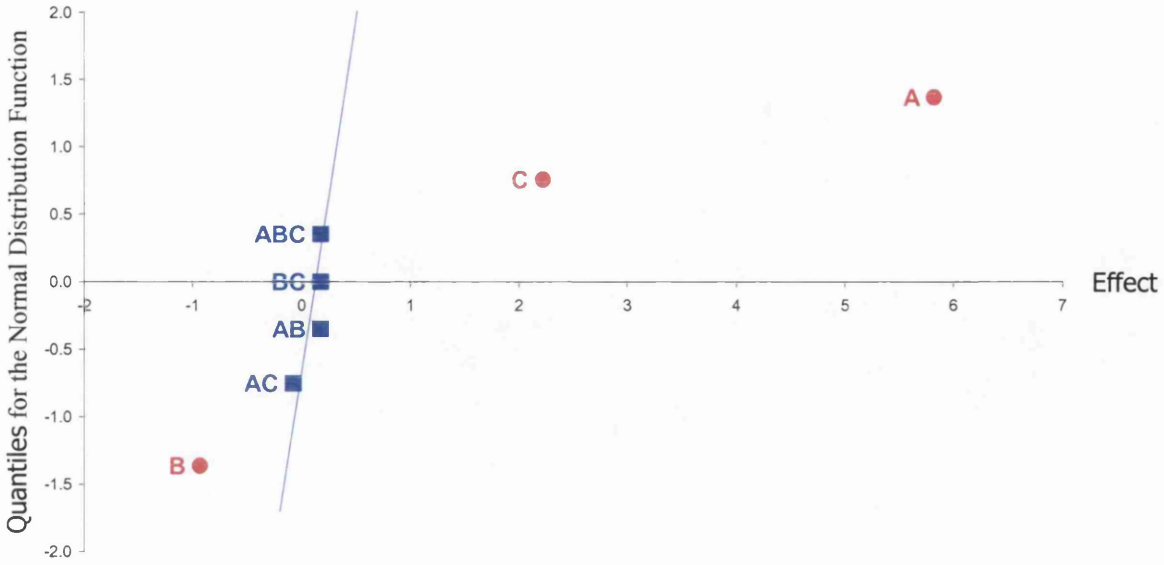


Figure 12 Normal probability plot for the experiment results.

Effects A, B & C, highlighted in red, do not seem to be the result of background noise as they are off the blue line (Figure 12). This graphical technique is useful for identifying candidates for inclusion in a model, i.e. the influencing factors or interactions. ANOVA (explained in section 4.1.2) is needed to justify selection and to quantify the significance of any selected terms. Having determined these terms are significantly influencing the response Y over and above background experimental noise levels, they can be legitimately included in an equation. This equation, or model, can be used to predict the behaviour of the system and therefore allow optimisation. For this DoE, the equation is linear. Designs of higher resolution allow the addition of quadratic and/ or cubic terms to be added to the linear equation giving a more descriptive model. These extra non-linear terms are available at a cost of extra runs required. The equation for this DoE, using the significant terms A, B and C is: -

$$Y = I + C_A A + C_B B + C_c C$$

Where Y is the estimate of the response, I is the average of the responses from each run, $C_A A$ is the coefficient for factor A multiplied by the value of factor A etc. The

coefficients are equal to half the effects, calculated in Table 2. Therefore the model for this DoE is:-

$$Y = 97.01 + 2.91A - 0.46B + 1.11C$$

Comparison of coefficients allows the relative influence of the factors to be assessed. For example factor B has the effect of reducing Y whilst factor A is over twice as influential as factor C. This model needs to be altered to be useful for inputting real values for the factors to estimate the response. This is because 1 factor unit is really some other number of units e.g. 1 unit of factor A is really 30units, i.e. 30°C.

4.1.2 Statistical Filters Used by DoE to Justify Decisions and Conclusions

4.1.2.1 ANOVA

Analysis of variance, henceforth ANOVA, is the main technique used by DoE to assign a probability to the validity of a DoE model and all its terms. It allows judgement to be made about how likely observed changes of DoE response can be attributed to changes in the levels of individual factors or combinations of factors. Generally ANOVA examines whether subgroups of results differ from the overall set of results by comparing the dispersion of the subgroup with the spread of the remaining data set.

ANOVA can therefore be applied to DoE as changing a factor from high to low levels produces a subgroup of data. Consider the DoE example represented in Table 2. The effect of changing levels of factor C is 2.2 units. An experimenter needs to know if this effect is a significantly large enough one not to be attributed to experimental noise. If this effect is significant, then the spread between the average high level and average low level would be significantly greater than the spread of the data set containing solely experimental noise. The technique for comparing spreads is the F-test which tests the probability that two spreads have come from the same population.

The F-test probability depends on the magnitude of difference between the two variances, or more precisely the ratio of the larger to the smaller, and the sample size. F tables are used to look up these significances.

The measure of spread used is the 'Mean Square' or MS. MS is calculated by dividing the 'Sum of Squares' by the 'Degrees of freedom' (the number of independent comparisons available to estimate this parameter) or:-

$$MS = \frac{SS}{DF}$$

$$SS = \frac{\text{Effect}^2}{\text{Adjustment Term}}$$

$$\text{Adjustment Term} = \frac{\text{sum of squared coefficients}}{\text{subgroup size}}$$

So for the example in Table 2, using data for factor C, the sum of squared coefficients can be found by replacing the '+'s and '-'s with –1's and 1's.

	Run	Factor C	Coefficient	Squared coefficients
Subgroup 1	1	-	1	1
	2	-	1	1
	3	-	1	1
	4	-	1	1
Subgroup 2	5	+	-1	1
	6	+	-1	1
	7	+	-1	1
	8	+	-1	1
		Totals	0	8

Table 3 Data for the adjustment term calculation for factor C

Hence:

$$\text{Adjustment Term} = \frac{8}{4}$$

therefore:

$$SS = \frac{8 \text{ x } 2.2^2}{4} = 9.9$$

And therefore for factor C:

$$MS = \frac{9.9}{1} = 9.9$$

This value is now compared with the MS of the data solely attributable to experimental noise (residual data), i.e. from the factor combinations AB, AC, BC and ABC, detected by the graph in Figure 12. This is done by looking up the ratio

$MS_C : MS_{Residuals}$

Where $MS_{Residuals} = (MS_{AB} + MS_{AC} + MS_{BC} + MS_{ABC})$

in F tables. These tables state that the probability of getting this ratio of spreads from data from the same population is less than 1/10,000. In other words it is statistically likely that factor C is influencing the response when changed from low to high level or vice versa. The ANOVA for the rest of the factors in the example DoE has been calculated and tabulated in Table 4. This ANOVA table is typical of the format used in subsequent chapters.

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	79.47	3	26.49	543.41	< 0.0001
A	67.86	1	67.86	1392.03	< 0.0001
B	1.71	1	1.71	35.10	0.0041
C	9.90	1	9.90	203.10	0.0001
Residual	0.195	4	0.05		
Cor Total	79.67	7			

Table 4 ANOVA table for the doe example

DF is the degrees of freedom. Degrees of freedom for the model is the number of model terms, including the intercept, less one. Degrees of freedom for a model term is the number of levels for the term, minus one. Degrees of freedom Cor Total is the number of terms in the model and not in the model, including the intercept less 1.

Degrees of freedom residual = DF Cor Total - DF Model.

There are several statistical indices that are used within a DoE report that indicate the validity of the model and therefore the believability of the predictions. Several are explained bellow, however their derivation is not explored.

4.1.2.2 Correlation Plots, R-squared and Adjusted R-squared.

The first index is obtained from a simple XY plot of the predicted values vs. the actual values. This provides a graphical way of assessing how well the model predicts the experiment data. Most of the data should lie close to a line of gradient 1 and intercept 0. Linear regression provides an estimate of the best-fit line's gradient and intercept which can be directly contrasted with the ideal line of gradient 1, intercept 0. The R-squared multiple correlation coefficient index or R^2 is a measure of the amount of variation around the mean explained by the model, calculated by

1-(SSresidual / (SSmodel + SSresidual))

where SSresidual is the sum of squares for all the terms not included in the model and SSmodel is the sum of squares of all the terms included in the model. The closer R^2 is to 1 the higher the correlation between the data. There is a related index called the adjusted R squared index. It is a measure of the amount of variation about the mean explained by the model using the R-squared multiple correlation coefficient adjusted for the spread of data. It is calculated by:-

1-((SSresidual / DFresidual) / ((SSmodel + SSresidual) / (DFmodel + DFresidual)))

It is essentially the equation for R-squared with DFresidual and DFmodel terms. DFresidual is the degrees of freedom of the residuals (terms not included in the model) and DFmodel is the degrees of freedom of terms included in the model. The closer to 1 the value, the higher the amount of variation explained by the model.

4.1.2.3 Adequate Precision

The adequate precision index is used to indicate the signal to noise ratio of the data. It compares the range of the predicted responses (i.e. the difference between the maximum response and the minimum response) at the design points with the average prediction error. Essentially it looks to see if there is enough discrimination in the model, thus allowing the modeller to use the model to make predictions within the model's scope without the noise swamping the signal. A value greater than 4 is desirable. It is calculated using the following equation:

$$\left[\frac{Max(Y) - Min(Y)}{\sqrt{V(Y)}}\right] > 4$$

Where $V(Y) = \frac{P\sigma^2}{n}$

Y = the estimated response or model prediction

P = number of model parameters including the intercept.

σ^2 = residual MS from ANOVA table

n = number of experiments

4.1.2.4 Lack-of-fit

Lack-of-fit is the variation of the data around the model. If the model does not fit the data well, this will be significant. Lack-of-fit can only be calculated when there are design points not included in the model. The model checks these points to see how well they are predicted. The difference between predicted and actual can be compared with the general error of the experiment and checked for significance. ANOVA is used to see if the lack-of-fit of this unused data is significant. This index is used in subsequent chapters where the DoE structures allow. It is desirable to have insignificant lack-of-fit.

4.2 Neural modelling

The relationship between variables and responses in complicated or new processes is usually not understood well enough to be able to write a logical statement or equation relating the two i.e. standard programming is not capable of solving this problem. This may be because of inherent non-linearity in the system, noise, highly interacted variables or unknown outliers in the data. An artificial neural network is able cope with these hurdles and therefore under these circumstances becomes the preferred tool over statistical approaches such as linear regression, DoE etc for deriving a model (providing there are sufficient legitimate examples of data from the system under investigation to model from). Neural models can also cope with large numbers of variables, scoring over statistical techniques such as DoE which require prohibitively large amounts of structured data when considering numbers of variables greater than about 7, especially if knowledge is required about curved relationships and multiple interactions.

4.2.1 Background

A neural network is a type of data processing architecture that is capable of analysing and making sense of complicated systems. The development of neurological models for brain function for artificial intelligence research was the original driver for the development of artificial neural network algorithms. Subsequent research in to their use in a variety of areas of problem solving has lead to various enhancements or modifications to these early models. This diversification has given rise to distinct sub-categories of artificial neural network algorithms.

Structurally, artificial neural networks consist of individual processing elements, called neurons, that receive data in the form of numbers from various connections. They combine the numbers, modify the result and pass that new number, or output, to subsequent neurons via connections.

Functionally they are non-linear data correlators that require no prior assumptions. Non-linear data correlators extract relationships (linear or non-linear) within the data. There are many types of data correlation technique for generating non-linear models, however most require some prior assumptions to be made and often have limitations that are too restrictive. They model by detecting potentially complex patterns in the data and are said to learn iteratively.

However, there are some pitfalls of using the neural network models. The neural network can learn the training data well and score very highly at predicting back the responses. However, it may not accurately predict responses for new examples that it has never seen before i.e. just because it has learned well doesn't mean that it can predict well. Several reasons may explain this discrepancy.

- **The present is not the past**. New examples may have been generated since the training data set and something has changed, e.g. a seasonal factor or new source of raw materials. The discrepancy means that there is an influential variable in the process not accounted for by the neural network. The corrective action would be to look for that variable and re-educate the network with the new data.
- **Biased data**. Consider a neural network that predicts simply a pass or failure for product from a process. If 99.9% of examples presented to the neural network in the training data set were examples of process conditions that gave passes (and therefore 0.1% that gave of failures) the result would be that the neural network would probably ignore that 0.1%. In predicting back the training data set, a score of 99.9% would be achieved. Failure would not be predicted too frequently to make the model useful. The corrective action for this would be to ensure that there are adequate numbers examples in the training data set to cover the area of interest.

- **Extrapolation outside of its experience**. The systems that neural networks are chosen to model are seldom linear therefore extrapolated predictions could be very inaccurate. Again the corrective action for this would be to ensure that there are adequate numbers examples in the training data set to cover the area of interest.
- **The neural network has sufficient complexity in it that it "over-learns the training data set**. Consider Figure 13.

Figure 13 Possible fits to noisy data taken from a smooth trend

The red line represents the true relationship. The training data approximates to this but due to natural variation/ noise some small deviations from theoretical were observed. A neural network with too little complexity, e.g. with no hidden layer, may approximate with a simple model, say a linear model, and therefore will under-fit. A neural network with too much complexity in the algorithm may over-fit. The green line represents over-fitting where every data point is exactly predicted. It has not generalised the noise and has not been vague enough to make accurate interpolations possible. The corrective action for this problem would be to test the resultant model with a fresh data set, not used to create the model, and assess its performance. Over-fitting would lead to removal of complexity from the algorithm and under-fitting would lead to adding complexity.

In summary a neural network model that highlights and describes complex relationships that are exclusive to the training data and do not exist globally (global in the sense of all the other possible data of interest to the modeller) are not useful i.e. they have modelled noise. Described relationships that do exist globally are useful and the neural network model is said to be able to generalise.

4.2.2 Back propagation neural network algorithm.

There are many types of artificial neural network. The type of neural network chosen to tackle this problem is the back propagation neural network. The back propagation neural network is a good generaliser. The software used to develop these back propagation neural network models was NeuralWorks Professional II/PLUS, version 5.30, by NeuralWare Inc. Generally back propagation neural networks take data, in the form of numerical values, as input signals to a neuron from one or more other connected neurons. The neuron assigns a numerical weighting to each input value, sums the resultant values and performs a mathematical function on that sum. This processed value is output to any subsequent neuron as its new input value. This part of the back propagation algorithm is illustrated in Figure 14 where W1 etc represent weights assigned to the inputs and f represents the mathematical function or transfer function.

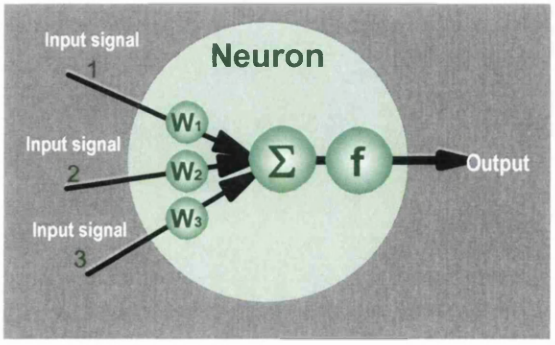


Figure 14 Artificial neuron, a mathematical algorithm for modifying inputs in to outputs.

The function, f, used in this investigation is known as the sigmoid transfer function. It is a function that converts the summation of weighted input values to a value of between 0 and 1. This function has the following equation[15]:-

$$T = (1 + e^{-S \times Gain})^{-1}$$

where

T = the result of applying the function to the current sum

S = the current sum of the weighted inputs

Gain = a slope altering constant.

These neurons are arranged in networks. The networks, for the back propagation neural network category, are typically arranged as follows. (Figure 15.)

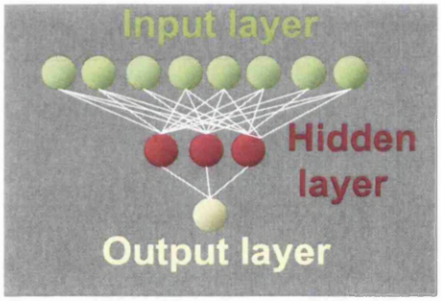


Figure 15 Back Propagation neural network architecture

There is an input layer, the layer that gets data into the network, one or more hidden or processing layers and an output layer. An output neuron of an output layer produces a number that can be translated in to a meaningful property, such as a measured quality parameter. These artificial neural networks have the ability to learn from experience or examples. A process that has several variables and produces an outcome or response, can be modelled or learned by an artificial neural network. Examples of the response and associated process variables (possibly including attributes such as colour, gender etc) are coded into numbers and fed in to the neural network. The data cascades through the network, which is initially configured

randomly, under the supervision of a learning algorithm (for supervised training neural networks). This algorithm minimises the difference between the real response data and the modelled response data at the output layer by adjusting weights, adjusting connections or effectively making and breaking connections etc until minimum error is achieved. In other words, repeated exposure to examples of the system under investigation allows the neural network to "formulate theories" as to how the system works. It then, using its latest theory suggests what the responses should be for a given set of process variables. If the guess is close to the response actually achieved in the real life example, the learning algorithm preserves most of the network and tweaks some aspects, however if the guess is way off, the learning algorithm is able to make more radical changes. This adjustment continues until the learning process is stopped or reaches some optimum. Features of these neural networks include adjustable parameters such as the learning coefficient (or learning rate term) and the momentum term. The learning algorithm used in this thesis is known as the Delta Rule. It works as follows. The error between the output layer prediction and the actual value is calculated to be the difference between the two. This error is then modified by the inverse of the sigmoid transfer function, back through the processing layer of neurons where it is collected as the error (e). This value, e, in conjunction with the momentum term and learning rate terms, modify each of the weightings for the inputs to the processing layer neurons. They do this using the following equation[15]:-

New Weighting = Old Weighting + Learning Rate * e * Input Value + Momentum Term * the Last Change of Weighting.

Hence the new weighting is in proportion to the error. The learning rate is the rate at which differences between predicted and actual response modify the weights. The momentum term ensures that if the weights change in a certain direction, then there is a tendency for weights to continue to change in that direction. The values for

momentum and learning rate terms were set at values previously experimented with during experience gained generating neural networks outside of the scope of this thesis[17]. The experimentation process was trial and error. No further validation was done on neural networks parameters in this thesis apart from making a couple of adjustments in the early stages and noting that the adjusted neural network algorithms yielded inferior models. This was not an exhaustive validation.

4.2.3 Neural Network Modelling Methodology Used

- **Generate the modelling data.** The data for modelling is selected from a master data set containing all data, comments etc. There are some important selection criteria for data selection for neural network modelling.
    a. There must be no very highly correlated input variables. Filtering for highly correlated input variables (i.e. the dependent variables that are associated) must result in discarding all but one representative input. For example, two adjacent extruder zone temperatures were recorded in the master database however due to the pragmatics of running the line, they were always changed together by similar amounts, therefore a single representative zone temperature is used. If in the future the master data set has data added where each is run at different temperatures such that the very high degree of correlation is reduced, then it may be valid to include the two as separate inputs.
    b. The data has to be numerical. Therefore something like the variable "machine scale" (i.e. pilot or full) must be numerically coded. In this case a 1 for pilot and a 2 for full suffices.
    c. Each example presented to the model for training or prediction must have the same number of inputs and outputs.
- **Split the data in to two data sets, training data and test data.** The training data set is used to develop the neural network model and the test set is used to estimate the model's predictive powers of new data.

- **Transform the data in to a format suitable for the neural network software.**
- **Specify the number of neurons in the hidden layer.** More or less neurons will increase or decrease the neural network's ability to over-fit or generalise. A guideline starting point suggested by the neural network software writers[15] for the number of hidden layer neurons should be equal to $\frac{t}{5(o+i)}$ where t is the number of training cases, o is the number of output neurons and i is the number of input neurons. There is no guarantee that this formula delivers an optimised network for the data. Therefore a method was developed for finding the optimum number of neurons in the hidden layer. The method is firstly to select the recommended number of hidden layer neurons using the equation above, train the model and note the RMS (root mean square) error of prediction for the test data set. The RMS error is an optional instrument that the software is able to calculate. Train the model again with more hidden layer neurons and again note the RMS error. Compare it with the previous value error. Use more or less hidden layer neurons in the next iteration depending on whether the addition improved or worsened the RMS error. Reiterate until optimum performance is achieved.
- **Define the number of cycles the neural network model trains over.** (This is done at the same time as finding out the optimum number of hidden layer processing elements). Being unaware of any rule of thumb that gives the optimum number of training cycles prior to training the neural network, a method was developed. The method is to set the maximum number of training cycles in the software to an excessively high number, say 100,000 and start the neural network learning. Every 3,000 training cycles the training data set (not the test data set) is predicted by the latest network configuration and the RMS error is noted. This is repeated until number of training cycles matches the maximum set in the software. This is repeated for each attempt at finding the

optimum number of hidden layer neurons described previously. The results are graphed as a plot of RMS error verses the number or training cycles for each network architecture as per the example in Figure 16. Typically there is an early rapid decrease in RMS error followed by a tailing off, similar to a plot of 1/x. Two visual judgements are then made. The first is when increasing the number of cycles fails to significantly improve the RMS error i.e. when the neural network is failing to learn anything new. The second is when the true hierarchy of network architectures has emerged. In the example in Figure 16, this approach would derive that the optimum number of cycles is about 75,000. Taking this superior architecture, which has the lowest RMS error, the training would be re-run, stopping after 75,000 cycles. In this example it would be the architecture represented by the green line.

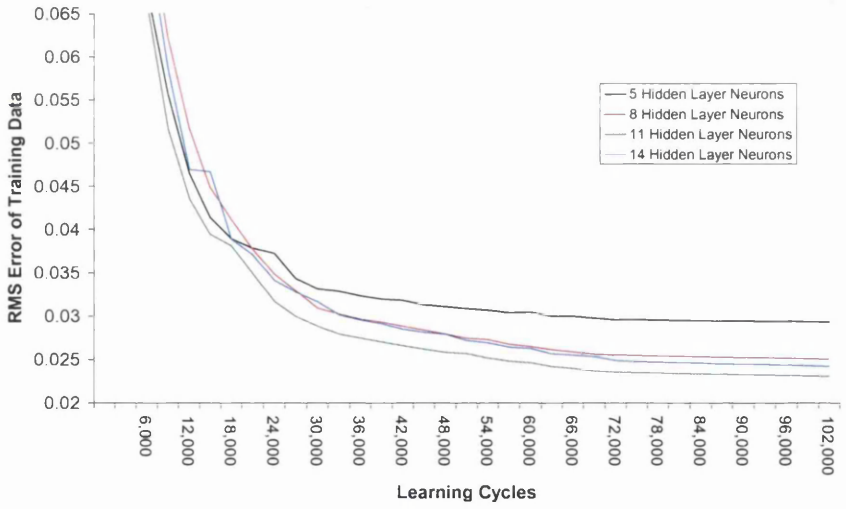


Figure 16 Graph of the RMS error vs training cycles for differning numbers of hidden layer neurons

- **Check the chosen model for over-fitting.** For this project, the method developed for doing this combines knowledge about the measurement system error with visually assessing the model's predictive performance. This is done by plotting the ranked predicted values of the training data set with the actual

values. Predictions would be expected to follow the underlying trend of the actual values if it had learned some true relationships, and at the same time the distribution of the residuals (actual–predicted) would be expected to match or exceed the distribution of normal test error. If this last criterion is matched, then an assumption can be made that the model has not inappropriately modelled noise.

- **Interrogate the neural network model and interpret the results.** One drawback of neural network models is that their equations are unwieldy and are extremely difficult to write out in high-level terms. It is easy enough to interrogate the model, telling it what levels of input variables you would like predictions for and getting a result but the software offers no contour mapping, main effects plots, interaction matrices or other visualisation tools common in statistical modelling techniques. Therefore the approach adopted to interrogate neural network models was to treat them as virtual processes, existing on a computer and to use DoE techniques to question them as if they were real processes. Because this is done in software on a virtual but representative process, with no materials or production time used, very elaborate and otherwise prohibitively expensive experiments can be run, stretching the limits of DoE's software. DoE software packages tend to limit themselves to practical numbers of variables and run lengths. With neural network models it is possible to operate at these limits and beyond. Given time, DoE software that can handle more variables than the commercial software can be written using spreadsheets. As time was a limited resource for this project, it was decided to work within the limitations of commercial software. This would minimise the chances of a programming error producing misleading results.

4.3 Closure

The statistical DoE methods and the neural network modelling technique used in this project have been described. Although the neural network was derived using proprietary software, there was a need to develop methods to evaluate and interrogate the complex models developed to ensure they were an accurate representation of the processes. In the next chapter, these methods are applied to develop a model of the pilot-scale system.

Chapter 5 Modelling the Pilot-Scale System

The pilot-scale equipment was used to generate a neural network model to reveal trends generic to the process, to highlight opportunities for improvement in the design of the production equipment at the design stage and as a first step towards the more comprehensive second generation neural network model. This preliminary neural network model is henceforth referred to as the "first generation neural network model" to differentiate it from the more detailed one generated later.

This chapter starts with a description of the data used for training the first generation neural network model and how DoE's were used to efficiently cover a large amount of variables and levels, followed by the application and results of the neural network modelling methodology laid out in Chapter 4. It concludes by detailing how the scale-up of the processes benefited from using this approach.

5.1 Description of the First Neural Network Model Data

The datasets for the first generation neural network model are summarised in Figure 17.

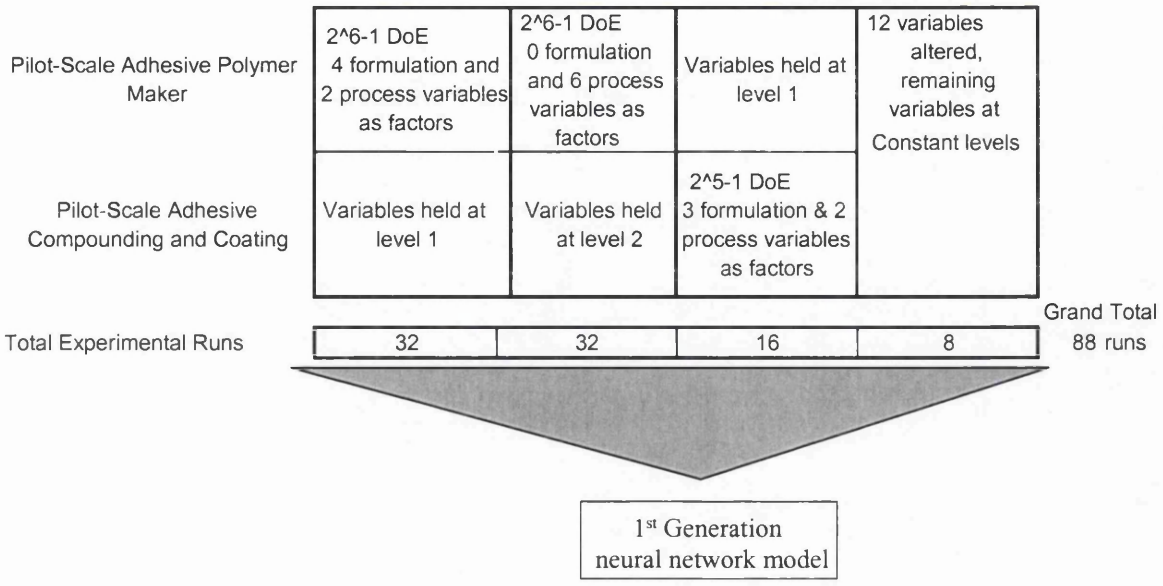


Figure 17 Summary description of data sets that fed the 1st generation neural network

Three DoE 's were run in total (detailed in appendix). Two were run on the pilot-scale adhesive polymer maker; each a 2^{6-1} fractional factorial design. One was run on the pilot-scale compounding and coating line, a 2^{5-1} fractional factorial design. Other process variables were altered between DoE's. In addition to the DoE's, data was collected for eight runs that included changes to 12 variables, unstructured by any formal design. These eight run conditions were set as part of adhesive polymer maker process equipment commissioning, an exercise outside of this planned research. Data from the runs was included in the neural network model because of the uniqueness of conditions each of these runs represented. The resulting adhesive polymer from each of the 88 experimental runs was made into tape and tested. The test results and process condition data were combined into a master database. No attempt was made to analyse these DoE's in their own right as DoE arrays were only used as a means of efficiently covering the large number of variables that could have contributed to tape properties. This approach makes use of the fact that DoE arrays are structured, i.e. have high and low levels so some thought could be applied to setting the levels of the factors; principally aiming to make the change significant enough to see an effect but small enough not to miss detail. Another advantage of this approach is that the DoE

structures ensure that every point is unique, and therefore of value for modelling. Both these features add up to data generation efficiency. Where DoE methodology falls short of being useful as a modelling tool in this situation is that it does not make provision for factors outside of the DoE to vary and still provide models with integrity. Therefore if DoE methodology was pursued, all the factors of interest would have to have been covered by a single DoE. If all the factors needing study were included in a low resolution DoE, more than 33,500,000 experimental run would have been required. Alternatively the number of factors reduced. Therefore to overcome these drawbacks, a neural network was used to make sense of the data. This data violated DoE rules by purposely altering variables (as well as experiment factors) both between and within the DoE's. As previously discussed in Chapter 4, Section 4.2, neural networks are adept at modelling with this sort of data. Because of this, a 33,500,000 run exercise was reduced to an 88 run exercise.

5.1.1 Application of Neural Network Modelling Methodology as per the Theory and the Exploitation of the Results

The model was generated from the data matrix using the methodology specified in Chapter 4 and discussed below step by step.

Step 1. Generate the modelling data from the total data collected. All the data generated was tested and filtered for the following:

a. Highly correlated input variables. Filtering for highly correlated input variables results in discarding all but one representative input. For example, UV lamp power for each row of lamps was varied by proportional amounts at the same time for convenience. After filtering, the power of only one row of lamps' is represented in the data used for modelling. The levels of the discarded variables must be remembered during model interpretation as they are implied, not specified by the neural network.

b. The categorical data was converted to numerical data. For example the variable "aerobic/ anaerobic" was coded as 1 and 2 respectively.

It was decided to combine the responses of adhesion and rolling wheel tack in to one response, using a variable column in the data set to instruct the neural network which response value came from which test, i.e. this variable column contained a code of 1's or 2's depending on the origin of the response value. A variable column was also added for which side of the tape the result came from; liner-side or face-side. In effect, four responses face side adhesion, face side tack, liner side adhesion and liner side tack were converted in to one response. This means that for neural network training purposes, there was only one response. After considering the alternatives, i.e. one neural network predicting four responses or four neural network models each predicting one response, it was deemed simpler and more efficient to go with the 'one response' option.

Step 2. Split the data into two data sets, training data and test data. The data was not sub-divided into training data and test data sets. This was because of the symmetry and efficiency of DoE's used to generate the data in the first place and the limited amount of data available. The DoE's had ensured that there was no data that did not represent a boundary of the system needing to be modelled, therefore each data point was necessary for teaching the neural network about that region of design space; each data point was sufficiently different from the rest that it gives unique insight into the processes. Therefore all the data was used to train the model. Adopting this alternative modelling strategy exposed the risk that the model may over-fit the data. This is discussed later.

Step 3. Specify the number of neurons in the hidden layer and define the number of cycles the neural network trains over. The maximum number of training cycles was set in the software to an excessively high number, i.e. 102,000, and started the neural network learning. Every 3,000 training cycles the training data set was

predicted by the latest network configuration and the RMS error was noted. This was repeated until the number of training cycles matched the maximum set in the software. This was repeated for each attempt at finding the optimum number of hidden layer neurons. The results were graphed as a plot of RMS error verses the number or training cycles for each network architecture and is shown in Figure 18. Typically there is an early rapid decrease in RMS error followed by a tailing off in the rate of reduction.

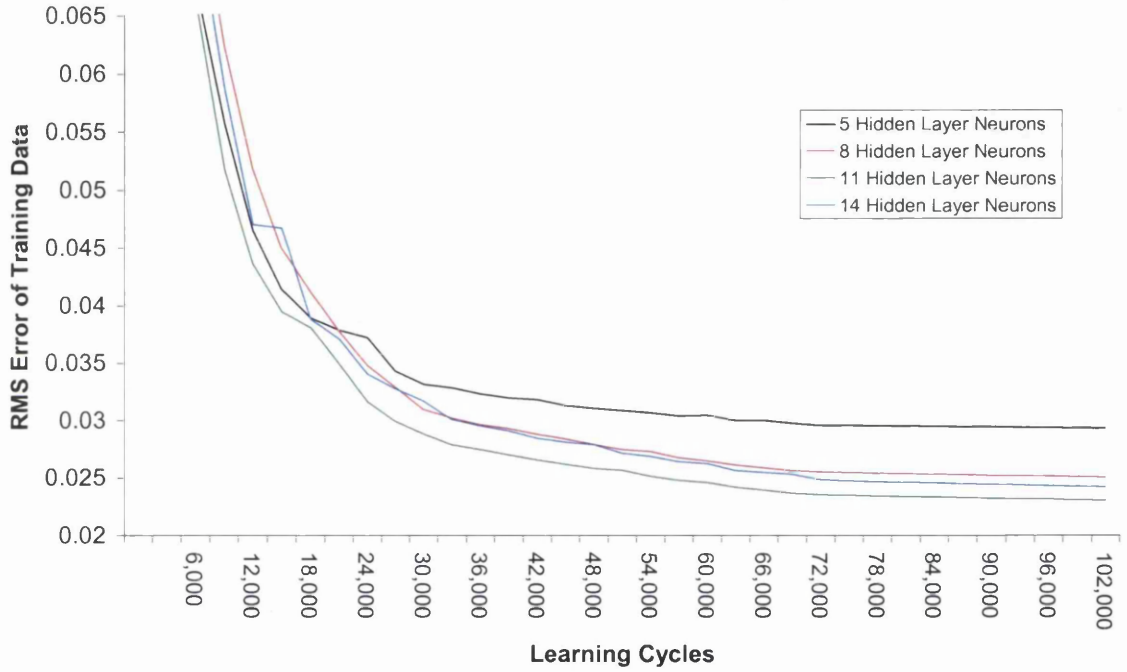


Figure 18 plot of RMS error vs Learning cycles for different numbers of hidden layer neurons

The graph above indicates that the optimum number of hidden layer neurons is 11 (because the lowest RMS error was achieved with this network configuration) and that about 72,000 learning cycles is sufficient to differentiate the RMS errors of the various neural network models (and therefore their accuracies). Those RMS errors remain relatively constant above 72,000 cycles.

The neural network with 11 hidden layer neurons was reset to its initial "ignorant" state and set to learn for 72,000 cycles only. The resulting model was checked for over-fitting.

Step 4. Check the chosen model for over-fitting. The predictions of the neural network model were compared with the actual values by plotting the data in Figure 19. These graphs are plots of the model's predictions for adhesion and tack verses the actual measured values, with a line of best-fit derived from linear regression. These graphs are intended to show two things. The first is if there is general agreement between the actual and predicted values. If there is, there will be a correlation between the data, the adjusted R^2 values will be close to 1, the intercepts will be close to 0 and the gradients close to 1. The second is to raise suspicion if the model has over-learned the training data, i.e. the neural network has modelled noise. This would be highlighted if the adjusted R^2 values were extremely close to 1 and, by eye, there was very little deviation from the line of best fit. A quantitative assessment would then be required to confirm if the prediction error is less than the historical test/ retest error for that measurement system.

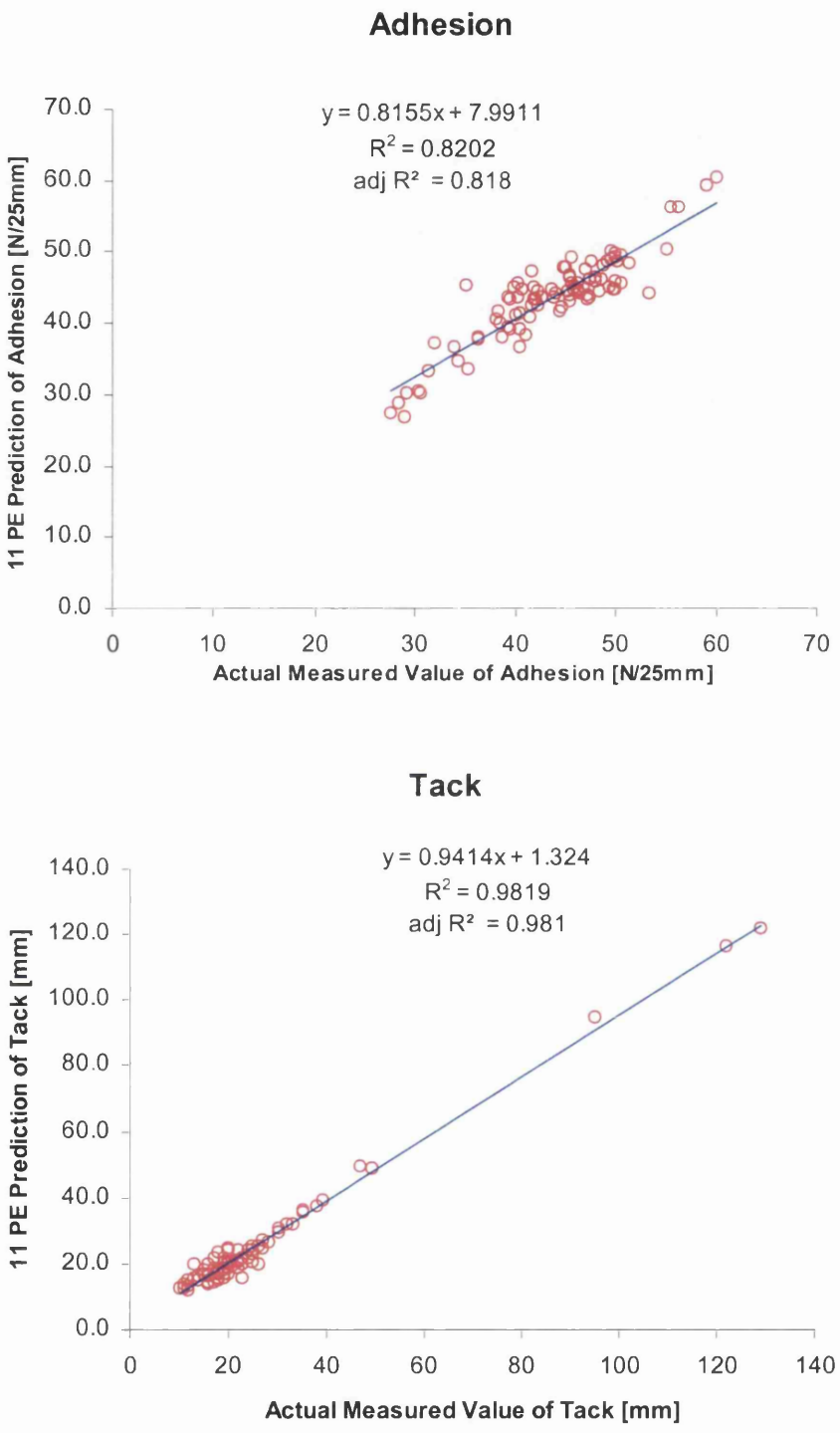


Figure 19 Correlation plots of predicted and actual measurements for Adhesion and rolling wheel tack.

The model's predictions of adhesion and tack seem to follow the underlying trend of the actual values. This is backed up by the ANOVA for each regression shown in Table 5. Therefore the neural network has learned some true relationships between the variables and responses, although there is some error in the prediction. The neural network has therefore not over-learned the data and has not over-modelled noise. It has been able to generalise.

ADHESION

Source	Sum of Squares	DF	Mean Square	F Value	Prob >F
Regression	2760.45	1	2760.45	411.08	0
Residual	604.36	90	6.72		
Total	3364.81	91			

ROLLING-WHEEL TACK

Source	Sum of Squares	DF	Mean Square	F Value	Prob >F
Regression	27574.9	1	27574.9	4774.43	0
Residual	508.2	88	5.8		
Total	28083.1	89			

Table 5 ANOVA for the linear regression of predicted vs actual responses

Step 5. Interrogate the neural network model and interpret the results.

The primary purpose for generating this neural network model of the pilot processes was to highlight the key parameters that affect finished product performance and assume that the full-scale processes are similarly affected. If the quality of the product is sensitive to the influence of the generic process and those influences lack adequate control then ongoing quality problems are likely to be endemic in any future scaled-up process. If this neural network model of the pilot processes can highlight these danger areas, then suitable engineering solutions can be included in the full-scale designs.

The way chosen to interrogate the first generation neural network model, to establish if changes to the designs of the full-scale production processes were required, was to run experiments on this virtual process. The neural network model represents virtual

pilot-scale processes whose variables can be set to any value within the scope of the neural network model's knowledge and the model will predict the product quality characteristics. These experiments were chosen specifically to highlight as many of the main effects as possible. For this a Taguchi L54 experimental design, a type of fractional factorial DoE, was chosen.

The Taguchi L54 design can analyse up to 26 variables, most for linearity. This design ignores interactions (that the neural network model is able to predict) and runs 25 of 26 at three levels (hence an indication of linearity). The remainder is run at 2 levels.

There was a relatively large number of variables that needed screening for their effects. Detail, such as interactions, was of no particular interest when designing the full-scale processes, therefore the Taguchi L54 design was appropriate. Linearity of the responses was of interest as this may have suggested if any theoretical areas of operation were more stable than others. This is why the Taguchi L54 was the design of choice.

It was decided to concentrate on 22 process variables and include 3 dummies (phantom variables). The inclusion of the dummy variables gave a measure of the error in the Taguchi experiment and help identify the significant variables. Any true effect would have more influence than a random one. Any apparent effects that were of similar magnitude to the dummy variables' effects could therefore be considered unimportant. The Taguchi L54 array is shown in Table 6 followed by table giving the uncoded factor levels. Level 1's are shown in yellow, level 2's in green and level 3's in lilac. The predicted adhesion and tack results are shown in the last two columns respectively. The variable codes correspond to those in Table 8 and Table 9.

ame	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y		
able ame	Side of tape tested f/s=1 l/s=2	Gas Flow [Vol per unit time]	Mass of A used in formulation	Antioxidant pph of total monomers	B:C used in formulation	A:C used in formulation	D:C used in formulation	Heating °	Agitation	Radiation Received	Reaction time	Average radiation source intensity	Minibatch size	Minibatch mass	adhesive age at time of coating	Coater uv Dose	tackifier1 : adhesive polymer	tackifier2 : adhesive polymer	additive: adhesive polymer	Extruder rpm / kg/hr	Time under UV s	average total thickness of coating	dummy 1	dummy 2	dummy 3	adhesion	rolling wheel
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	10.6	12.7
	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	12.3	16.3
	1	1	1	1	1	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	22.7	21.4
	1	1	2	2	2	2	2	2	1	1	1	1	1	1	2	3	2	3	2	3	2	3	2	3	2	19.6	33.6
	1	1	2	2	2	2	2	2	2	2	2	2	2	2	3	1	3	1	3	1	3	1	3	1	3	37.0	20.6
	1	1	2	2	2	2	2	2	3	3	3	3	3	3	1	2	1	2	1	2	1	2	1	2	1	23.0	21.4
	1	1	3	3	3	3	3	3	1	1	1	1	1	1	3	2	3	2	3	2	3	2	3	2	3	98.9	89.8
	1	1	3	3	3	3	3	3	2	2	2	2	2	2	1	3	1	3	1	3	1	3	1	3	1	85.6	87.4
	1	1	3	3	3	3	3	3	3	3	3	3	3	3	2	1	2	1	2	1	2	1	2	1	2	63.3	98.9
	1	2	1	1	2	2	3	3	1	1	2	2	3	3	1	1	1	1	2	3	2	3	3	2	3	12.3	12.4
	1	2	1	1	2	2	3	3	2	2	3	3	1	1	2	2	2	2	3	1	3	1	1	3	1	39.9	21.1
	1	2	1	1	2	2	3	3	3	3	1	1	2	2	3	3	3	3	1	2	1	2	2	1	2	33.5	22.1
	1	2	2	2	3	3	1	1	1	1	2	2	3	3	2	3	2	3	3	2	3	2	1	1	1	18.5	15.6
	1	2	2	2	3	3	1	1	2	2	3	3	1	1	3	1	3	1	1	3	1	3	2	2	2	12.6	19.3
	1	2	2	2	3	3	1	1	3	3	1	1	2	2	1	2	1	2	2	1	2	1	3	3	3	11.5	14.8
	1	2	3	3	1	1	2	2	1	1	2	2	3	3	3	2	3	2	1	1	1	1	2	3	2	56.5	40.1
	1	2	3	3	1	1	2	2	2	2	3	3	1	1	1	3	1	3	2	2	2	2	3	1	3	93.5	90.5
	1	2	3	3	1	1	2	2	3	3	1	1	2	2	2	1	2	1	3	3	3	3	1	2	1	57.2	46.0
	1	3	1	2	1	3	2	3	1	2	1	3	2	3	1	1	2	3	1	1	3	2	2	3	3	33.2	39.7
	1	3	1	2	1	3	2	3	2	3	2	1	3	1	2	2	3	1	2	2	1	3	3	1	1	20.4	15.8
	1	3	1	2	1	3	2	3	3	1	3	2	1	2	3	3	1	2	3	3	2	1	1	2	2	13.8	28.3
	1	3	2	3	2	1	3	1	1	2	1	3	2	3	2	3	3	2	2	3	1	1	3	2	1	90.2	92.6
	1	3	2	3	2	1	3	1	2	3	2	1	3	1	3	1	1	3	3	1	2	2	1	3	2	23.7	20.6
	1	3	2	3	2	1	3	1	3	1	3	2	1	2	1	2	2	1	1	2	3	3	2	1	3	43.5	30.5
	1	3	3	1	3	2	1	2	1	2	1	3	2	3	3	2	1	1	3	2	2	3	1	1	2	11.7	12.3
	1	3	3	1	3	2	1	2	2	3	2	1	3	1	1	3	2	2	1	3	3	1	2	2	3	19.4	22.0
	1	3	3	1	3	2	1	2	3	1	3	2	1	2	2	1	3	3	2	1	1	2	3	3	1	14.1	23.8
	2	1	1	3	3	2	2	1	1	3	3	2	2	1	1	1	3	2	3	2	2	3	2	3	1	37.7	47.0
	2	1	1	3	3	2	2	1	2	1	1	3	3	2	2	2	1	3	1	3	3	1	3	1	2	21.0	21.4
	2	1	1	3	3	2	2	1	3	2	2	1	1	3	3	3	2	1	2	1	1	2	1	2	3	26.1	12.9
	2	1	2	1	1	3	3	2	1	3	3	2	2	1	2	3	1	1	1	1	3	2	3	2	2	31.7	45.5
	2	1	2	1	1	3	3	2	2	1	1	3	3	2	3	1	2	2	2	2	1	3	1	3	3	41.1	55.7
	2	1	2	1	1	3	3	2	3	2	2	1	1	3	1	2	3	3	3	3	2	1	2	1	1	103.0	118.9
	2	1	3	2	2	1	1	3	1	3	3	2	2	1	3	2	2	3	2	3	1	1	1	1	3	63.0	59.2
	2	1	3	2	2	1	1	3	2	1	1	3	3	2	1	3	3	1	3	1	2	2	2	2	1	48.5	45.5
	2	1	3	2	2	1	1	3	3	2	2	1	1	3	2	1	1	2	1	2	3	3	3	3	2	15.5	13.3
	2	2	1	2	3	1	3	2	1	2	3	1	3	2	1	1	2	3	3	2	1	1	3	2	2	34.4	26.0
	2	2	1	2	3	1	3	2	2	3	1	2	1	3	2	2	3	1	1	3	2	2	1	3	3	34.8	32.5
	2	2	1	2	3	1	3	2	3	1	2	3	2	1	3	3	1	2	2	1	3	3	2	1	1	39.1	23.5
	2	2	2	3	1	2	1	3	1	2	3	1	3	2	2	3	3	2	1	1	2	3	1	1	3	48.1	34.0
	2	2	2	3	1	2	1	3	2	3	1	2	1	3	3	1	1	3	2	2	3	1	2	2	1	33.1	35.1
	2	2	2	3	1	2	1	3	3	1	2	3	2	1	1	2	2	1	3	3	1	2	3	3	2	29.1	37.9
	2	2	3	1	2	3	2	1	1	2	3	1	3	2	3	2	1	1	2	3	3	2	2	3	1	29.8	19.7
	2	2	3	1	2	3	2	1	2	3	1	2	1	3	1	3	2	2	3	1	1	3	3	1	2	33.3	20.9
	2	2	3	1	2	3	2	1	3	1	2	3	2	1	2	1	3	3	1	2	2	1	1	2	3	49.2	99.1
	2	3	1	3	2	3	1	2	1	3	2	3	1	2	1	1	3	2	2	3	3	2	1	1	2	50.4	57.8
	2	3	1	3	2	3	1	2	2	1	3	1	2	3	2	2	1	3	3	1	1	3	2	2	3	12.6	12.5
	2	3	1	3	2	3	1	2	3	2	1	2	3	1	3	3	2	1	1	2	2	1	3	3	1	28.5	17.8
	2	3	2	1	3	1	2	3	1	3	2	3	1	2	2	3	1	1	3	2	1	1	2	3	3	26.7	14.0
	2	3	2	1	3	1	2	3	2	1	3	1	2	3	3	1	2	2	1	3	2	2	3	1	1	15.2	12.8
	2	3	2	1	3	1	2	3	3	2	1	2	3	1	1	2	3	3	2	1	3	3	1	2	2	58.9	45.5
	2	3	3	2	1	2	3	1	1	3	2	3	1	2	3	2	2	3	1	1	2	3	3	2	1	65.2	108.4
	2	3	3	2	1	2	3	1	2	1	3	1	2	3	1	3	3	1	2	2	3	1	1	3	2	61.5	101.3
	2	3	3	2	1	2	3	1	3	2	1	2	3	1	2	1	1	2	3	3	1	2	2	1	3	66.6	87.6

Table 6 Taguchi L54 array.

Factor	Code	Level 1	Level 2	Level 3
Side of tape tested f/s=1 l/s=2	A	1	2	
Gas Flow [Vol per unit time]	B	0	20	40
Mass of A used in formulation	C	4.7	8.6	12.5
Antioxidant pph of total monomers	D	0.3	0.5	0.7
B:C used in formulation	E	64.95	99.98	135.01
A:C used in formulation	F	8.87	16.445	24.02
D:C used in formulation	G	1.85	4.15	6.45
Heating °	H	60	72.5	85
Agitation	I	0	80.5	161
Radiation Received	J	1960.945	2373.473	2786
Reaction time	K	567	731	895
Average radiation source intensity	L	2.191	3.356	4.521
Minibatch size	M	39.5	60.25	81
Minibatch mass	N	296.6	367.95	439.3
adhesive age at time of coating	O	113	128	143
Coater uv Dose	P	600	800	1000
tackifier1 : adhesive polymer	Q	25	35	45
tackifier2 : adhesive polymer	R	5	15	25
additive: adhesive polymer	S	0.3	0.5	0.7
Extruder rpm / kg/hr	T	25.9	33.6	41.3
Time under UV s	U	1.45	1.88	2.31
average total thickness of coating	V	220	294.5	369
Dummy1	W	1	2	3
Dummy2	X	1	2	3
Dummy3	Y	1	2	3

Table 7 Factor names, name codes and uncoded factor levels for the L54 Taguchi DoE

Each row of the L54 matrix was presented to the first generation neural network model and predictions of adhesion and tack made. This data was analysed as a Taguchi L54 and the effects for the different levels calculated. The differences (deltas) between the level that gave the maximum effect and that which gave the minimum effect for each variable was calculated and ranked. ANOVA was performed on the results and summarised in Table 8 and Table 9. The P values from these ANOVA's are summarised in Figure 20. They show the ranked P values in order of decreasing significance for factors A-Y for both adhesion and tack. The red line at P= 0.1 represents the chosen cut-off for significance. Values >0.1, or above the red line were deemed to be indistinguishable from noise. The justification for selecting the P = 0.1 significance level is the dummy variables, W,X & Y, that are a measure of noise, therefore anything with similar P values or greater was considered to be noise. Anything with a P value just under that of the dummy variable, Y (the most significant dummy variable for adhesion) was counted as significant. This was

because the consequences of reacting to a false influencing factor was deemed preferable to ignoring a true one that the Taguchi failed to highlight. Especially when it is remembered that Taguchi DoE's are crude modelling tools. Normal probability plots were also done for each response, the results of which are shown in Figure 21 and. Figure 22. The significant effects are highlighted in red. These were done to check agreement of significance using a different approach.

Factor	Mean effect of adhesion at level 1	Mean effect of adhesion at level 2	Mean effect of adhesion at level 3	Delta (max-min)	nth largest effect
G	25.7	36.2	53.8	28.12	1
C	26.8	37.0	51.8	24.98	2
D	31.4	34.3	49.9	18.52	3
Q	31.2	35.7	48.8	17.56	4
J	33.6	45.9	36.1	12.33	5
R	32.5	39.6	43.6	11.10	6
Y	42.2	32.4	41.1	9.82	7
E	44.4	36.3	35.0	9.44	8
F	41.5	33.3	41.0	8.19	11
O	43.9	35.6	36.2	8.21	10
X	42.9	38.8	34.0	8.89	9
P	33.7	40.8	41.1	7.40	12
A	36.2	40.9		4.66	21
V	42.4	37.9	35.4	6.96	13
L	35.6	37.8	42.2	6.62	14
S	34.8	41.1	39.7	6.27	15
B	42.3	37.0	36.4	5.83	16
H	35.3	39.3	41.0	5.71	17
M	40.8	39.5	35.4	5.44	18
T	36.3	37.9	41.5	5.17	19
I	41.0	35.9	38.8	5.16	20
N	41.2	37.4	37.0	4.21	22
K	40.7	38.3	36.7	3.99	23
W	39.8	38.4	37.5	2.37	24
U	37.8	38.4	39.5	1.75	25

	Sum of Squares	DF	Mean Square	F Value	Prob > F
	7262.6	2	3631.3	27.5	0.005
	5676.4	2	2838.2	21.5	0.007
	3572.7	2	1786.3	13.53	0.017
	2997.5	2	1498.7	11.35	0.022
	1524.9	2	762.5	5.77	0.066
	1133.5	2	566.7	4.29	0.101
	1044.6	2	522.3	3.96	0.113
	941.8	2	470.9	3.57	0.129
	759	2	379.5	2.87	0.168
	757.9	2	379	2.87	0.169
	715.1	2	357.6	2.71	0.18
	631.5	2	315.8	2.39	0.207
	292.1	1	292.1	2.21	0.211
	448.1	2	224.1	1.7	0.293
	410.1	2	205	1.55	0.317
	392.4	2	196.2	1.49	0.329
	370.7	2	185.3	1.4	0.345
	309.7	2	154.8	1.17	0.397
	290.6	2	145.3	1.1	0.416
	250.7	2	125.4	0.95	0.46
	241.9	2	120.9	0.92	0.47
	195	2	97.5	0.74	0.533
	144.7	2	72.3	0.55	0.616
	51.3	2	25.6	0.19	0.831
	28.7	2	14.4	0.11	0.9
Residual	528.1	4	132		
Total	30971.7	53			

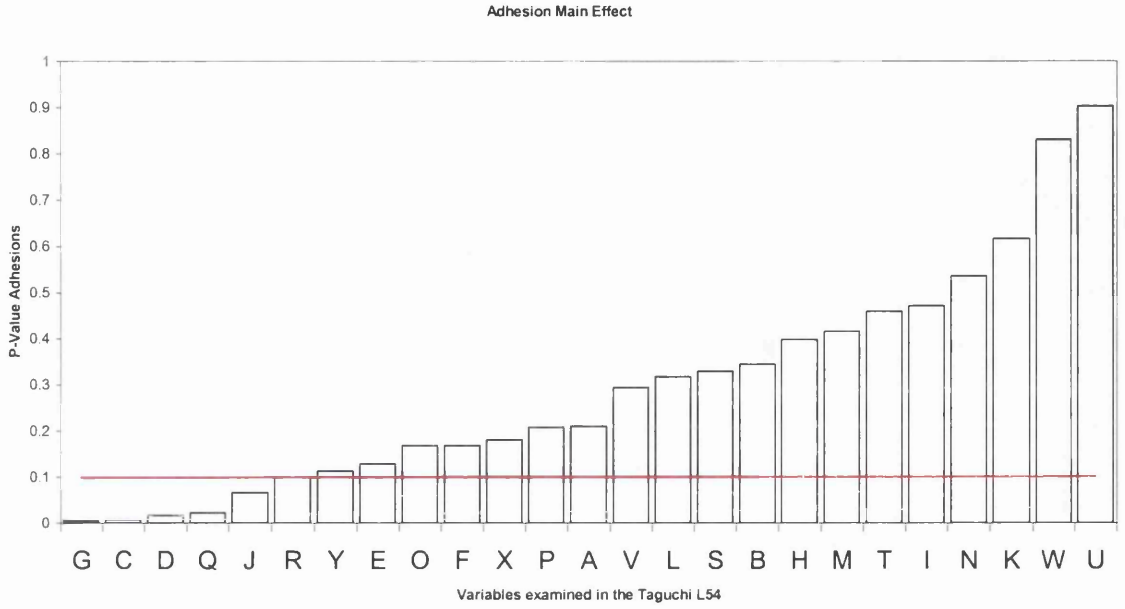
Table 8. Analysis of Adhesion showing the effect at each level, the delta, its rank and the ANOVA.

Factor	Mean effect of R-W Tack at level 1	Mean effect of R-W Tack at level 2	Mean effect of R-W Tack at level 3	Delta (max-min)	nth largest effect
C	24.5	37.1	59.2	34.69	1
G	26.2	34.9	59.7	33.53	2
E	51.9	34.5	34.4	17.53	4
Q	31.8	37.5	51.5	19.67	3
R	33.1	38.8	48.9	15.84	5
D	33.1	39.2	48.5	15.36	6
L	34.9	37.0	48.9	13.99	7
A	35.9	44.6		8.74	15
F	35.7	37.5	47.6	11.92	8
B	45.6	33.9	41.3	11.71	10
O	46.2	40.2	34.4	11.80	9
V	46.9	38.4	35.5	11.35	11
U	36.8	46.9	37.2	10.10	12
M	42.4	43.9	34.5	9.41	14
T	35.6	40.2	45.0	9.43	13
X	41.9	43.2	35.6	7.61	16
S	37.8	45.0	38.1	7.19	17
I	42.2	36.2	42.4	6.24	19
N	43.8	37.3	39.7	6.46	18
Y	42.5	36.7	41.6	5.78	20
J	38.7	43.1	39.1	4.39	21
W	42.4	40.3	38.1	4.31	22
H	42.3	38.9	39.6	3.42	23
K	40.3	40.9	39.6	1.31	24
P	40.5	39.9	40.5	0.59	25

	Sum of Squares	DF	Mean Square	F Value	Prob > F
	11105	2	5552.5	23.03	0.006
	10882.3	2	5441.2	22.56	0.007
	3677.4	2	1838.7	7.62	0.043
	3689.9	2	1845	7.65	0.043
	2316.3	2	1158.2	4.8	0.086
	2154.5	2	1077.2	4.47	0.096
	2050.2	2	1025.1	4.25	0.102
	1033.2	1	1033.2	4.28	0.107
	1488.3	2	744.2	3.09	0.155
	1263	2	631.5	2.62	0.187
	1255.7	2	627.9	2.6	0.189
	1252	2	626	2.6	0.189
	1173.2	2	586.6	2.43	0.204
	917.3	2	458.6	1.9	0.263
	801.1	2	400.5	1.66	0.298
	595.4	2	297.7	1.23	0.382
	592.7	2	296.4	1.23	0.384
	454.6	2	227.3	0.94	0.462
	384.3	2	192.2	0.8	0.511
	348.7	2	174.4	0.72	0.539
	212.7	2	106.4	0.44	0.671
	167.8	2	83.9	0.35	0.726
	118.1	2	59	0.24	0.794
	15.2	2	7.6	0.03	0.969
	4.1	2	2.1	0.01	0.991
Residual	964.6	4	241.1		
Total	48917.8	53			

Table 9. Analysis of Rolling Wheel Tack showing the effect at each level, the delta, its rank and the ANOVA.

Adhesion



Rolling Wheel Tack

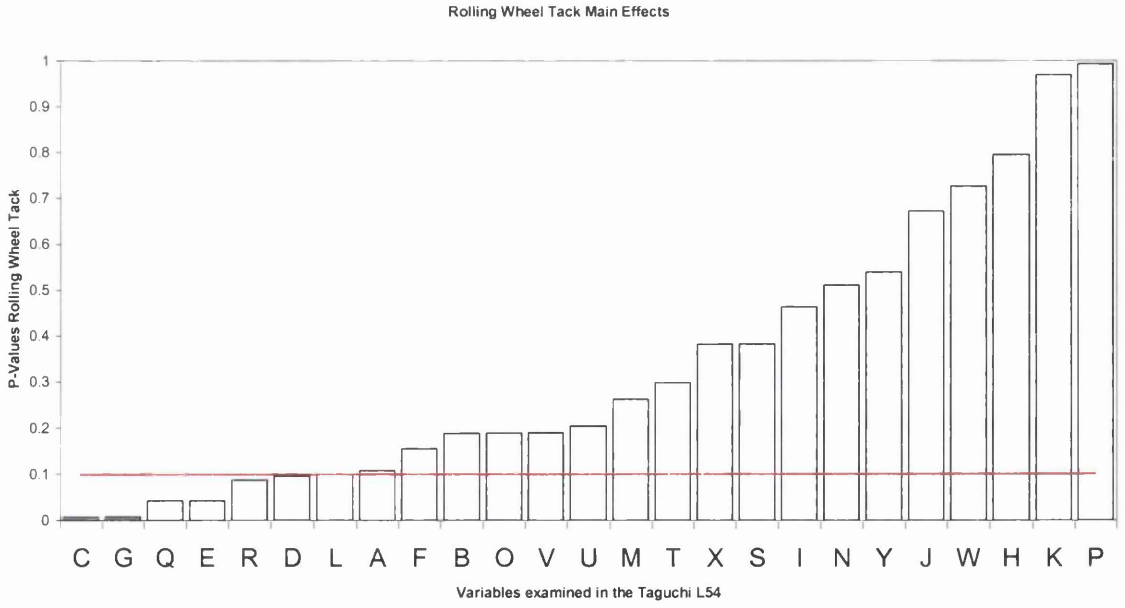


Figure 20 Ranked P values in order of decreasing significance for factors A-Y for adhesion and tack

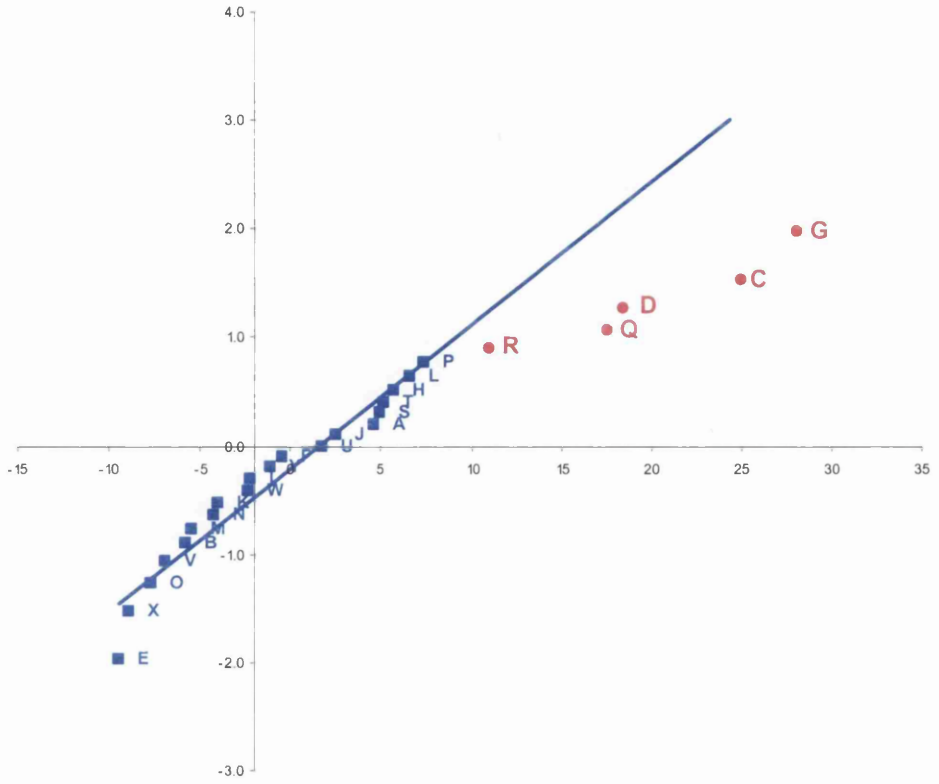


Figure 21 Normal probability plot for Adhesion.

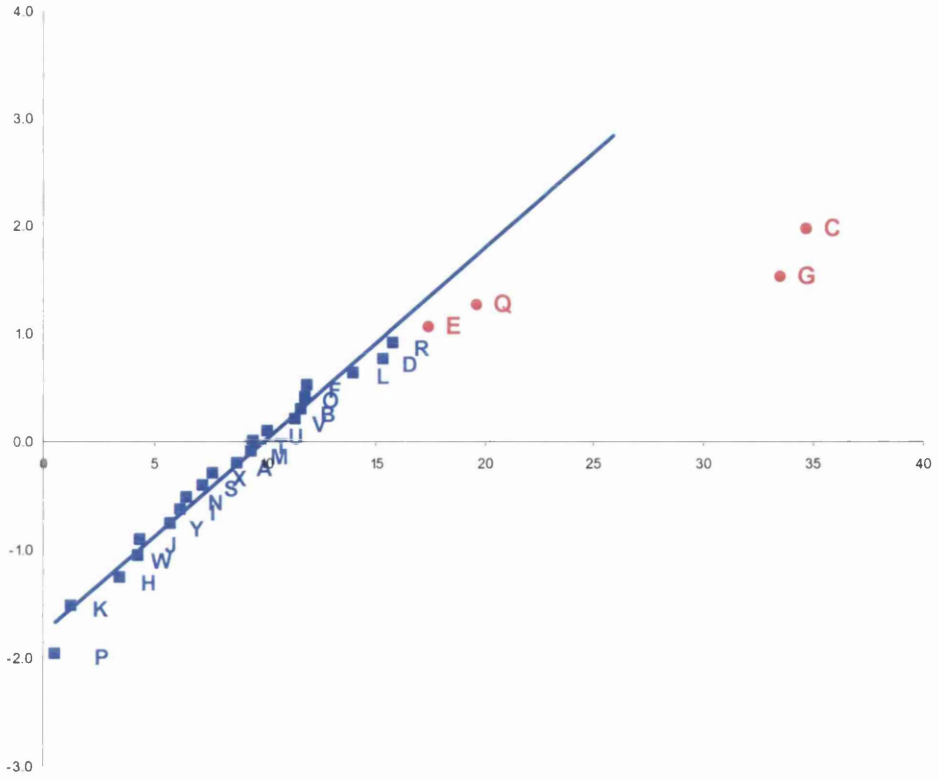


Figure 22 Normal probability plot for Rolling Wheel Tack.

The significant effects (i.e. variables with the largest deltas, that are statistically significant with p-values less than 0.1 and that stand out on the normal probability plots) are formulation-related and radiation related, both on the pilot-scale adhesive making process and on the pilot-scale compounding and coating process.

To check the significant main effects for linearity, the effects at levels 1-3 were plotted for each response in Figure 23. These plots are designed to give an appreciation of the affect on the responses of changing a factor from level 1 to 2 to 3. Factors G,C,D & Q effect both adhesion and tack. J shows a maximum at level 2, but only affects adhesions. E falls to an apparent plateau, but only affects tack. The rest have an approximately linear affect, increasing with level.

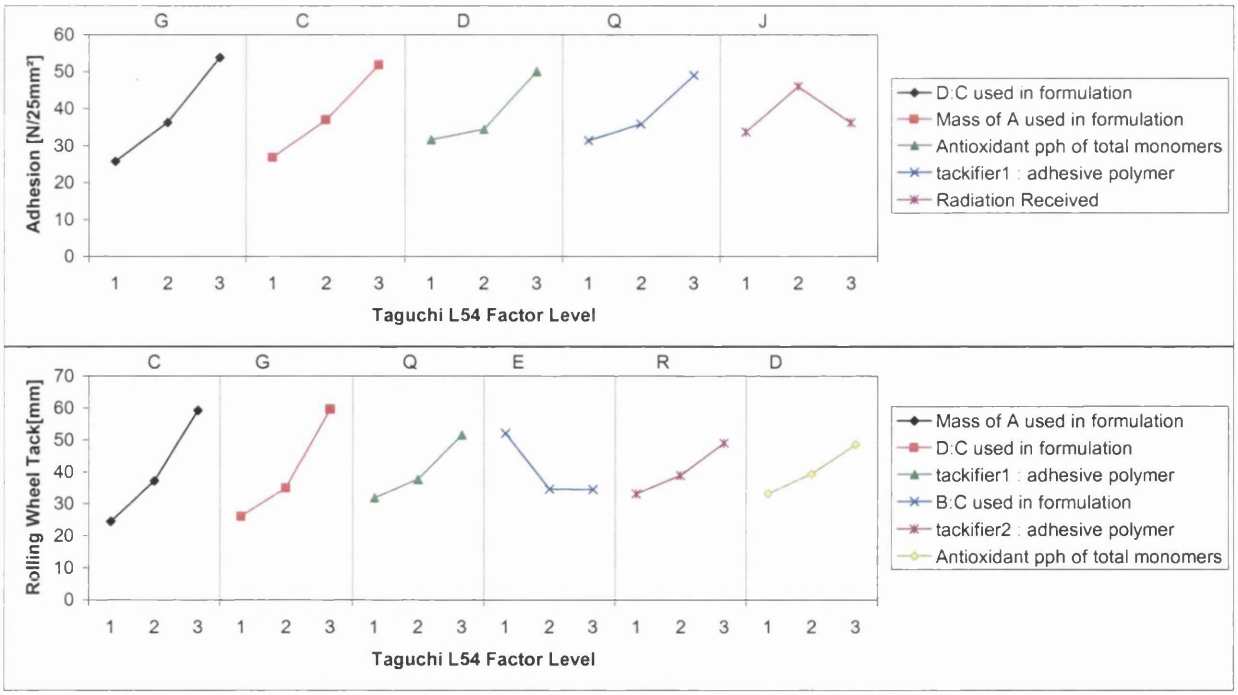


Figure 23 Main effects plots for the significant Factors in the Taguchi L54 interrogation of the first generation neural network mode.

5.1.2 Impact of the findings on the final process design

Taguchi DoE's are low-resolution designs that are not ideal for modelling complex interactions or complex curvature. Therefore it is expected that this Taguchi DoE does not come close to representing the complexity of the first generation neural network model. The Taguchi found some factors that were significant and failed to prove the significance of others. This does not mean that the insignificant factors have no effect, only that either their effect is too small to stand out from the background noise, or that the levels chosen were too spread out to find small areas of interest. With these caveats, several points of relevance to the design and optimisation of the future processes emerged.

1. For adhesion, the optimum "radiation received" variable setting lies between levels 1 and 3, therefore extra radiation capacity in the future process may not be needed to get the most efficient results. See Figure 23.
2. Some significant factors appear in both the adhesion and tack graphs, ie "mass of A used in formulation", "antioxidant pph of total monomers" and "tackifier1: adhesive polymer". Using these factors alone to increase the adhesion will also increase the rolling wheel tack results. (The slopes of the graphs are in the same direction). This means that if those factors are used to control adhesion, an increase in the adhesion leads to a reduction of tack (the further the wheel travels the lower the tack).
3. Simplistically, for the plots in Figure 23 that show linearity, the implication is that by extrapolating even further, higher tack or higher adhesion is possible. Future experimentation should therefore explore past the limits of the current neural network model in an attempt to explore "super-properties".

5.2 Closure

Using a series of DoE's on the different processes to generate data for a neural network model has been successful. Insights into the pilot-scale processes have been

gained. The formulation and radiation aspects of the processes are important for consistency and quality. The success of this modelling phase meant the designs for the full-scale processes were revisited to ensure reliable, accurate and precise metering technologies for the raw materials were chosen and that the most consistent UV sources were purchased.

The modelling of the full-scale compounding and coating sub-processes are described in the next chapter.

Chapter 6 Modelling the Full-Scale Compounding and Coating Sub-Processes

Phase 1 of the project was designed to provide information that would help design a robust full-scale process, good at controlling variables critical to tape quality. It involved developing and learning from the first generation neural network model, and did indeed influence the type of systems procured for full-scale processes.

The full-scale compounding-and-coating process was established before the full-scale adhesive polymer maker. Therefore it was possible to generate data for the second generation neural network model with adhesive from the pilot-scale adhesive polymer maker but converted to tape on the full-scale compounding-and-coating processes. This is referred to as phase 2.

The elements of the full-scale compounding-and-coating processes were commissioned in stages, and therefore became available at different times for studying. These elements were characterised and optimised independently, each treated as a sub-process.

Each sub-process outputs a response that becomes a variable for the full process optimisation later, e.g. the amount of adhesive polymer in the formulation, a variable used in the full-scale process optimisation, is a result of various polymer feeder variables acting and interacting together to produce a delivery rate response. Another example is that the dose of UV radiation given to the coated tape, a variable in the full-scale process optimisation but a response in the sub-process characterisation. UV Dose is a result of the variables of line speed, lamp power and the number of lamps in some combination. The sub-systems that were characterised are as follows:

1. **The adhesive-polymer-metering system.** Characterisation of this system is important because the first generation neural network model showed that the ratio of the tackifying resins to adhesive polymer is a major influence on adhesion and rolling wheel tack.
2. **The tackifying resin and additive addition system.** This also contributes to the same main effects that makes the adhesive-polymer-metering system important, i.e. the ratio of tackifying resins to adhesive polymer.
3. **The UV curing system.** The output is of this sub-process is the variable "dose of UV radiation within a specific band of wavelengths". This variable is a known main effect for a tape quality parameter that was not modelled by the first generation neural network model. This quality parameter is "shear strength". Shear strength is tested by hanging a weight on a splice made from the tape, and timing to when the splice fails. This is treated as a pass/ fail attribute as, because of the way it is tested, it is difficult to get a continuous scale of meaningful numbers. UV dose was assumed to be scaleable from a lab-scale unit. This lab-scale unit influenced the design of the full-scale UV curing system. The assumptions in the scaling up to the full-scale process needed to be validated by characterising the full-scale UV curing system and comparing with the earlier findings.

This chapter deals with a description of each sub-process and the application of statistical modelling techniques to characterise them. The purpose of the characterisation is to be able to confidently set the outputs of the sub-processes as variables in future experimentation. These systems have been highlighted as critical for consistent and high tape quality by the first generation neural network model.

6.1 The full-scale adhesive-polymer-metering system.

6.1.1 System Description

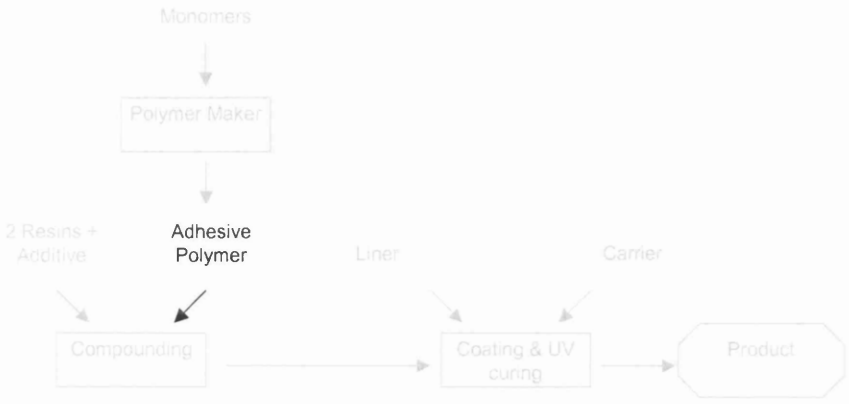


Figure 24 Map of the generic complete process highlighting the polymer-metering process

The adhesive-polymer-metering system is the means by which adhesive polymer is continuously fed into the twin-screw extruder in a controlled manor. The profitability of the compounding-and-coating process requires a minimum throughput to be attained. The diagram below details the system.

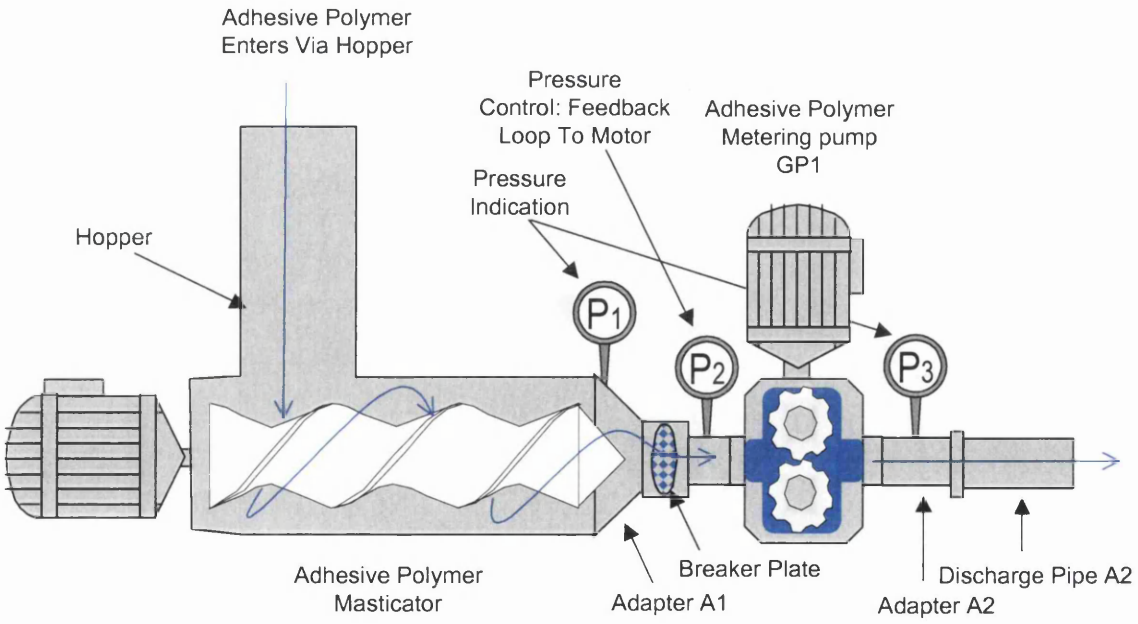


Figure 25 Schematic of the adhesive-polymer-metering system.

The adhesive polymer enters the system via the masticator hopper. The masticator, a crude single screw extruder, masticates the adhesive polymer by the screw action within a diminishing clearance barrel. The adhesive polymer is extruded out of the masticator, into adapter A1 and therefore pressurises the inlet to GP1. The pressure is maintained at set-point by speed modulation of the masticator screw.

GP1 is a volumetric metering gear pump that meters the adhesive polymer. Each revolution of the gears moves a fixed volume. The mass of adhesive polymer pumped therefore depends on how much material fills and leaves the gears each revolution. GP1 pumps adhesive polymer through A2 and the discharge pipe into the twin-screw extruder.

The system can be heated by controlling several independent heating zones that are illustrated in Figure 26.

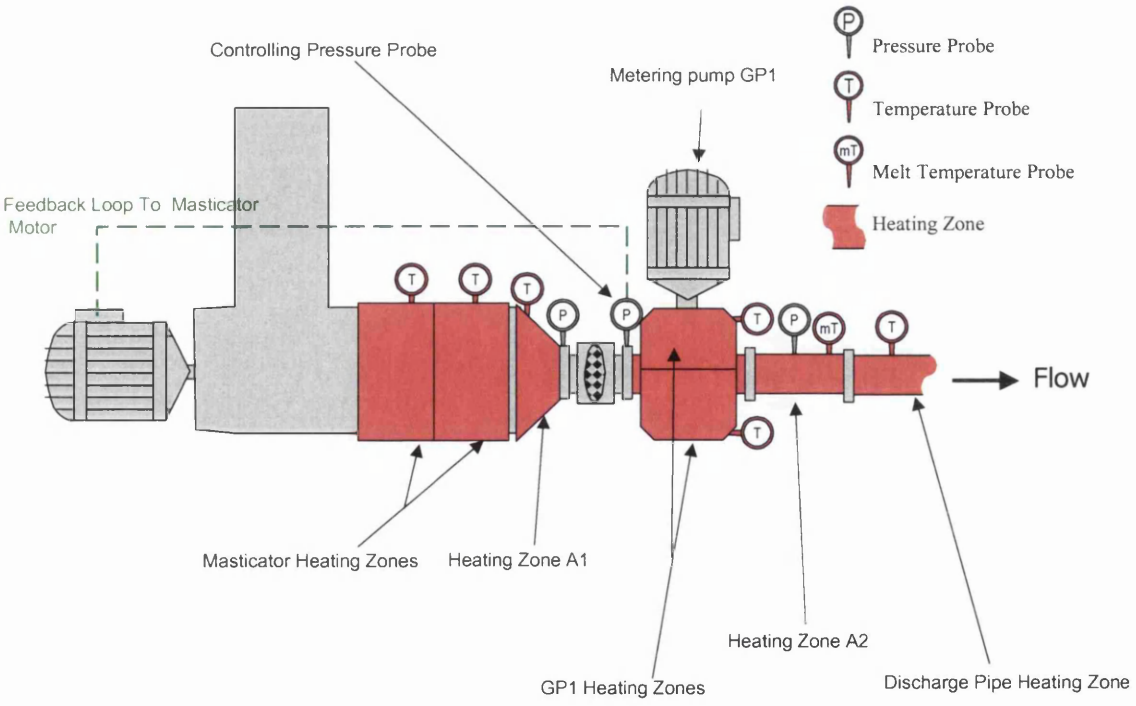


Figure 26 Adhesive-polymer-metering system heater arangements, temperature sensing points and pressure sensing and control points.

The masticator heating zones consists of heaters around the barrel of the masticator. The heater zone A1, includes heaters around adapter A1. The GP1 heater zone includes upper and lower heating plates on GP1. The heater zone A2 includes heaters around adapter A2 and around the discharge pipe. Each zone is closed loop controlled using thermocouple feedback. Temperature set-points are controlled by the SCADA system (Supervisory Control And Data Acquisition). A deviation of measured temperature from set-point is counteracted by a proportional increase or decrease of time that the heater is on using a PID controller. For example if the thermocouple is indicating that the system is well below temperature, the PID controller will instruct the heater(s) to stay on for relatively long periods of time. When the thermocouple senses that the system is approaching the set-point temperature, the heater(s) switch off and on in shorter bursts. Each heater zone has one thermocouple but can have several heater elements. The thermocouples only measure the temperature of the metal near the thermocouple. Therefore within each heating zone there could be a range of temperatures. The melt will experience each of these temperatures within a zone and receive a non uniform but a controlled average heat history.

The pressure at the inlet to GP1 is closed loop controlled. Pressure is increased or decreased by speeding up or slowing down the masticator screw.

The proprietary nature of the polymer meant some assumptions had to be made about equipment sizing and metering capability, particularly for the masticator which is used in the food industry. These assumptions had to be proven during commissioning.

This system is volumetric i.e. the metering pump, a positive displacement gear pump, transfers a constant cavity per revolution. If the amount of adhesive polymer in the cavity is unaffected by process variables then the system will simply be to calibrate to deliver a constant mass of adhesive polymer. If the amount of adhesive polymer in the cavity is affected by process variables then the system will be less simple to calibrate

to deliver a constant mass of adhesive polymer as the mass will depend on those process variables. The philosophy of the whole compounding system requires this sub-process to deliver a target mass of adhesive polymer per unit time. Therefore the system relies on a volumetric device to achieve gravimetric delivery. The objective of the research is to calibrate the volumetric system to be a gravimetric system. There is no on-line measurement of throughput so this needs to be modelled under all process conditions. The major goal of modelling this sub-process was to gain the control and understanding of the process variables that influence throughput. It was necessary to optimise these variables and be able to adjust throughput according to the requirements of manufacturing. In this investigation throughput is a response but in later process optimisation experiments, throughput is a factor. A robust model needs to be derived for throughput, and the throughput variability. NB Costs associated with running equipment hotter, faster or harder were not considered as optimisation constraints.

The objective of the study of this sub-process was to establish:-

1. Whether it was possible to get an accurate model for throughput, thereby translating a volumetric device into a gravimetric device
2. Whether it was possible to minimise throughput variability
3. What was the temperature of the melt exiting the feeding system and therefore entering the extruder.

6.1.2 Investigation study

Preliminary investigation study experiments were run in order to establish the extreme operating limits of the metering system, thus allowing the setting of runable limits for a DoE. These experiments included running at extremes of temperature and extremes of inlet pressures. They suggested that the system was far from linear and therefore a central composite DoE, that has good curvature modelling properties, was chosen as

the most appropriate modelling method. Central composite DoE's, because of their requirements for factors to be run at multiple levels, are able to model curvature with quadratic or cubic formulae. The fact that they can introduce curvature into the models using quadratic or cubic equations means that relatively complicated relationships can be modelled. A central composite DoE using four factors was chosen as the most appropriate design to answer the research questions.

The factors for the experiment are listed in Figure 27. Factor A, the pre-pump heating zones temperature set-point, is a group of individual temperature zones that condition the adhesive polymer entering GP1. Factor B, the GP1, A2 and discharge pipe heating zones temperature set-point, is a group of individual temperature zones that condition the adhesive polymer within the pump and after the pump. Factor C, GP1 speed, moves any material within the pump through the system. Factor D is the pressure of the adhesive polymer entering metering pump. This pressure is controlled by the speed of the masticator (under closed loop control). It conditions the flow force of the adhesive polymer at the inlet of the pump.

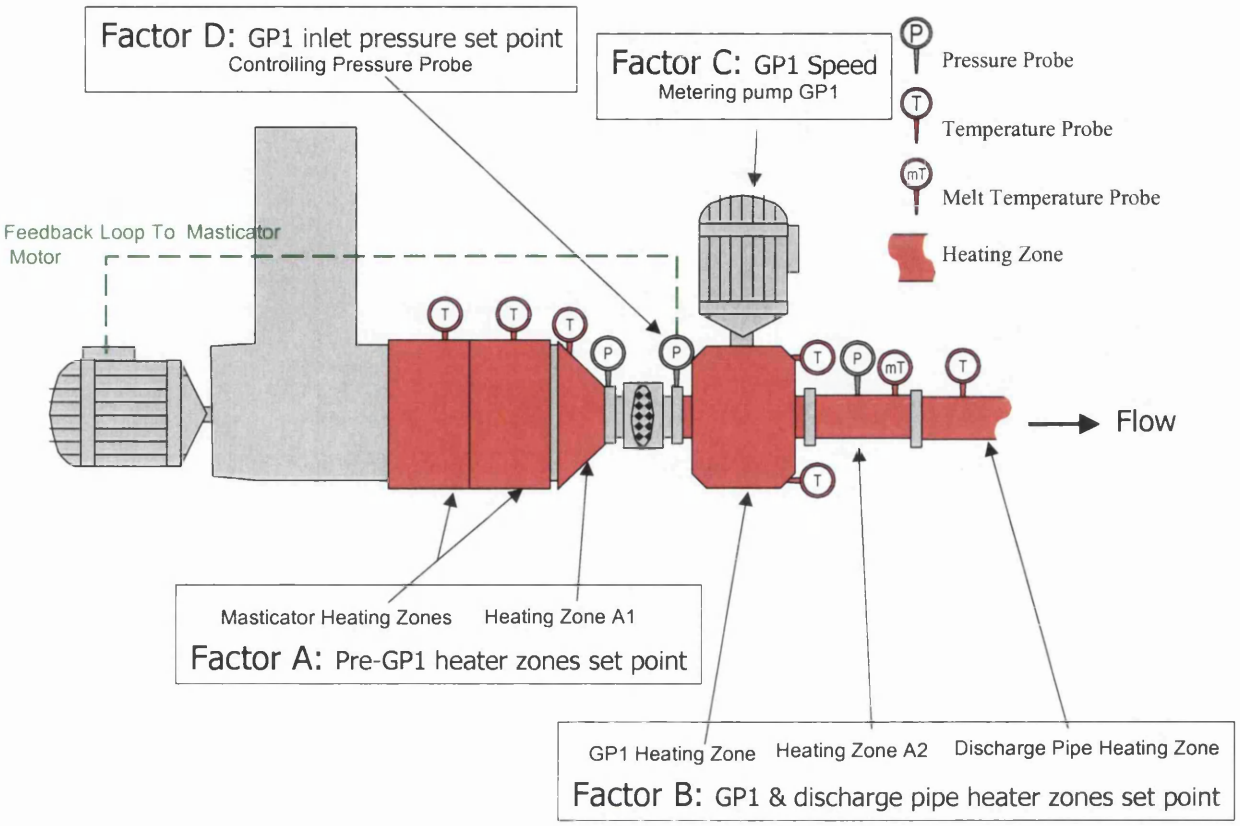


Figure 27 Schematic of the adhesive-polymer-metering system with some details about the factors used to generate data for the central composite DoE and the temperature and pressure measurement points.

All relevant data is collected in real time by the SCADA. The responses are delivery rate, delivery rate variation and adhesive melt temperature. The delivery rate was measured by collecting and weighing a number of timed outputs from the system for each experimental condition. The mean weight of each experimental condition was reported as the throughput. The delivery rate variation was measured by calculating the standard deviation of the repeat weighings. The melt temperature was taken using an infrared pyrometer pointed at the flow as it exited the discharge pipe. The DoE structure and results are shown in Table 10.

		Factors				Responses		
Standard Order	Run Order	A Pre-GP1 Temp sp [°C]	B GP1 & Discharge Pipe Temp sp [°C]	C GP1 Speed sp [rpm]	D GP1 Inlet Pressure sp [bar]	Throughput [kg/hr]	Throughput standard deviation [g/min]	Melt Temp °C
1	3	120	120	46.25	19.5	103.77	11.06	95.5
2	8	160	120	46.25	19.5	102.87	9.19	103
3	21	120	160	46.25	19.5	103.09	8.48	119
4	25	160	160	46.25	19.5	102.92	6.31	123
5	4	120	120	58.75	19.5	127.28	26.93	93.2
6	9	160	120	58.75	19.5	130.22	11.66	106.5
7	22	120	160	58.75	19.5	129.51	9.83	110
8	26	160	160	58.75	19.5	130.88	5.34	122.5
9	5	120	120	46.25	22.5	103.23	7.14	94.75
10	10	160	120	46.25	22.5	102.33	4.98	109
11	23	120	160	46.25	22.5	103.60	10.35	111
12	27	160	160	46.25	22.5	102.32	5.04	126
13	6	120	120	58.75	22.5	130.68	6.89	98.25
14	11	160	120	58.75	22.5	129.72	8.26	112
15	24	120	160	58.75	22.5	131.26	9.57	112
16	28	160	160	58.75	22.5	131.01	6.97	125
17	13	100	140	52.5	21	3.92	99.37	108
18	20	180	140	52.5	21	115.75	2.18	120
19	1	140	100	52.5	21	116.83	4.99	94.5
20	29	140	180	52.5	21	117.17	6.59	136
21	14	140	140	40	21	89.48	6.49	112
22	16	140	140	65	21	144.51	11.91	109
23	17	140	140	52.5	18	117.01	6.55	110.5
24	18	140	140	52.5	24	117.33	5.64	112.5
25	2	140	140	52.5	21	117.50	10.69	111
26	7	140	140	52.5	21	118.31	21.52	107.75
27	12	140	140	52.5	21	117.49	10.17	113
28	15	140	140	52.5	21	117.59	7.25	108
29	19	140	140	52.5	21	116.81	1.85	115

Table 10 Central composite DoE factors, factor levels and responses

6.1.3 Analysis of the DoE : ANOVA

6.1.3.1 ANOVA for the Throughput Model and Interpretation

The ANOVA output for the throughput is given in Table 11.

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	4457.086	9	495.2318287	1638.806	< 0.0001
A	0.042101	1	0.042101258	0.13932	0.7133
B	1.107595	1	1.107594735	3.665218	0.0716
C	4442.174	1	4442.173993	14699.91	< 0.0001
D	0.753809	1	0.753808815	2.494481	0.1317
A^2	3.72913	1	3.729130089	12.34032	0.0025
AC	2.521426	1	2.52142641	8.343827	0.0098
AD	2.749793	1	2.749793062	9.09953	0.0074
BC	1.57741	1	1.577410403	5.219918	0.0347
CD	2.219951	1	2.219951003	7.346193	0.0143
Residual	5.439432	18	0.302190653		
Lack of Fit	4.297522	14	0.306965879	1.075272	0.5255
Pure Error	1.141909	4	0.285477363		
Cor Total	4462.526	27			

R-Squared	0.998781
Adj R-Squared	0.998172
Pred R-Squared	0.99652
Adeq Precision	165.6497

Table 11 ANOVA for the quadratic model predicting throughput

The F Value associated with the model is >1638. The F value is the ratio of the Model Mean Square to the Residual Mean Square and shows the relative contribution of the model variance to the residual variance. A large number indicates more of the variance being explained by the model; a small number indicates the variance may be due to noise. Prob > F is the probability of the observed F value if the null hypothesis is true, i.e. that there is no difference between the model mean square and the residual mean square. Small probabilities (less than 0.05) indicate that there is a model effect; large values (greater than say 0.10) suggest no significant effect. In this case the value

is <0.0001 and therefore the model effect is large. This value implies that the chances of getting an F value this large if the result was due to noise is less than 0.01%. Therefore the result is almost certainly due to altering the factors.

The F value for the lack of fit of the model is <1.1. This F value is the ratio of the Lack of Fit Mean Square to the Residual Mean Square. The higher this number, the greater the effect of this component of the residual. The Prob > F of the lack of fit shows the significance of this value. Probabilities > 0.1 = questionable significance while Probabilities < .05 = probably significant. It is desirable to have an insignificant lack of fit (P>0.1). In this case it is >0.52, therefore there is no significant lack of fit. This value implies that the chances of getting an F value this large if the result was due to noise is greater than 52%. Because there is no lack of fit, there is no observed behaviour in the experiment that is not adequately explained by the chosen model. If the lack of fit was significant then the model would have been unsuccessful at modelling certain observed behaviour during the experiment.

Each of the terms chosen for inclusion in the model, i.e. A,B,C,D,A^2,AC,AD,BC and CD, has an associated F value and Prob>F value. This F value is the ratio of the Mean Square for each of the terms to the Residual Mean Square. The Prob>F values show what the chances are of getting F values this large if the results were due to noise. All the chosen factors' F values are almost certainly due to altering the factors except possibly A,B and D. These are included because they interact significantly. For example they interact with C. The terms AC and BC are important to the model therefore, whilst there is no statistical reason not to include C, A and B must also be included to allow inclusion in the interaction terms. This justifies the inclusion of all the terms A,B,C,D,A^2,AC,AD,BC and CD in the model.

The adjusted R squared value is a measure of the amount of variation about the mean explained by the model using the R-squared multiple correlation coefficient adjusted

for the spread of data. The closer to 1 the value, the higher the amount of variation explained by the model. In this case the index is >0.998, therefore the model explains the majority of variation observed during the experiment.

The adequate precision index is a measure of signal to noise ratio. A value greater than 4 is desirable. In this case the value is >165, therefore indicating a discriminating model if used to investigate the territory covered by the experiment.

The robustness of the throughput model is shown graphically in Figure 28. Figure 28 is a plot showing the correlation between model predictions and actual results for throughput. The fact that the gradient is close to 1 and that the intercept is close to 0 means that there is very little disagreement between predicted and actual results. This is in line with the conclusions about the model drawn from the ANOVA output.

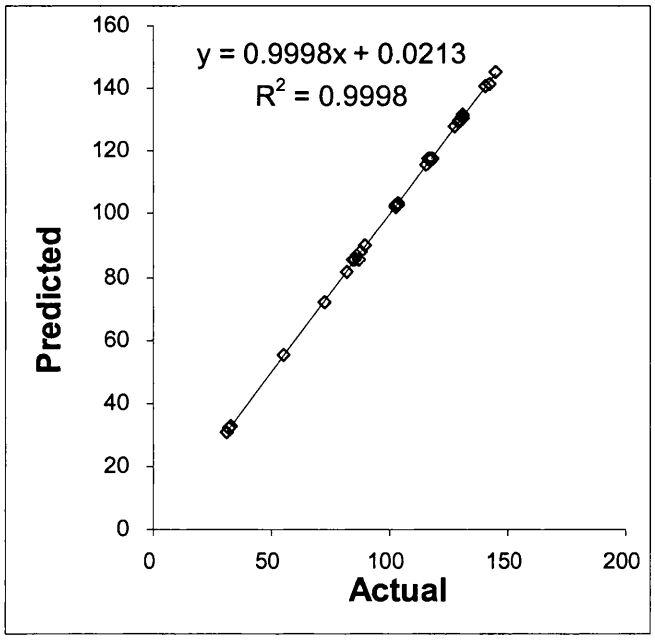


Figure 28. XY graph showing the correlation between predicted and actual throughputs based on the quadratic model.

6.1.3.2 ANOVA for the Melt Temperature Model and Interpretation

The ANOVA output for the melt temperature model is given in Table 12

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	2577.158	2	1288.57875	114.2155	< 0.0001
A	573.3038	1	573.30375	50.81579	< 0.0001
B	2003.854	1	2003.85375	177.6151	< 0.0001
Residual	293.332	26	11.28199934		
Lack of Fit	253.782	22	11.53554467	1.16668	0.4952
Pure Error	39.55	4	9.8875		
Cor Total	2870.489	28			

R-Squared	0.897811
Adj R-Squared	0.88995
Pred R-Squared	0.862746
Adeq Precision	33.83239

Table 12 ANOVA for the linear equation predicting melt temperature

The F Value associated with the model is >114. The Prob>F value is <0.0001 and therefore the model effect is large. This value implies that the chances of getting an F value this large if the result was due to noise is less than 0.01%. Therefore the result is almost certainly due to altering the factors.

The F value for the lack of fit of the model is >1.1. The Prob > F of the lack of fit is >0.49, therefore there is no significant lack of fit. This value implies that the chances of getting an F value this large if the result was due to noise is greater than 49%. Because there is no lack of fit, there is no observed behaviour in the experiment that is not adequately explained by the chosen model. If the lack of fit was significant then the model would have been unsuccessful at modelling certain observed behaviour during the experiment.

Each of the terms chosen for inclusion in the model, i.e. A & B, has an associated F value and Prob> value. The Prob> values show what the chances are of getting F

values this large if the results were due to noise. Both the chosen factors' F values are almost certainly due to altering the factors. This justifies their inclusion in the model.

The adjusted R squared value is >0.89, therefore the majority of variation is explained by the model.

The adequate precision index is >33, therefore indicating a discriminating model if used to investigate the territory covered by the experiment.

The robustness of the melt temperature model is shown in Figure 29. Figure 29 is a graph showing the correlation between model predictions and actual results for melt temperature. The fact that the gradient is close to 1 and that the intercept is close to 0 means that there is very little disagreement between predicted and actual results. There is however some minor disagreement that does not detract from the usefulness of the model but is reflected in the adjusted R square value in the ANOVA in Table 12.

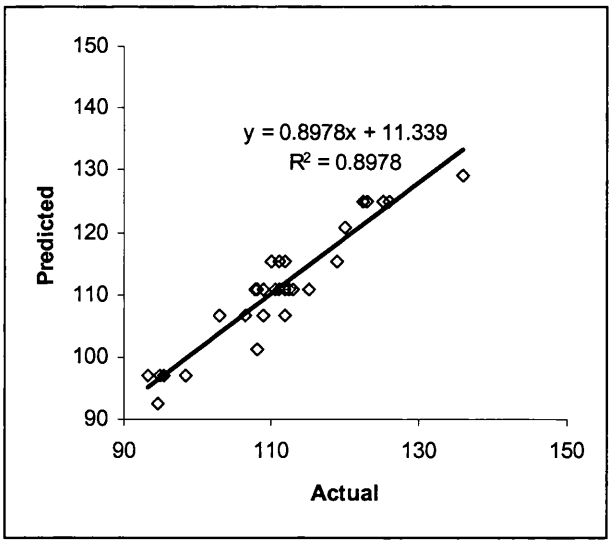


Figure 29 XY graph showing the correlation between predicted and actual melt temperatures based on the linear model.

6.1.4 Implications of the experimental study

It is possible to get a strong model for throughput provided the temperature of the adhesive polymer entering GP1 is above a threshold level. Section 6.1.3 shows that the throughput response is highly predictable using the quadratic equation model below:-

Throughput [kg/hr] = $19.52 + 0.4688 * A - 0.1211 * B + 0.5461 * C - 0.0332 * D - 0.00122 * A^2 + 0.00318 * AC - 0.01382 * AD + 0.00251 * BC + 0.03973 * CD$

Where

A = Pre-GP1 heater zones set point [°C]

B = GP1 & discharge pipe heater zones set point [°C]

C= GP1 Speed set point [rpm]

D= GP1 Inlet Pressure sp [bar]

As a confirmation, the system was set up on separate occasions to deliver a variety of throughputs. Factors A-D were set and the resulting throughputs measured. The model was asked to predict the throughputs for each set of conditions and these predictions and actual measurements are plotted (Figure 30). This graph demonstrates the correlation between the quadratic model and real life. The gradient of the correlation is close to 1 and the intercept close to 0. The R^2 is close to 1. All these indices imply that the accuracy of the model was confirmed over the full range of the system.

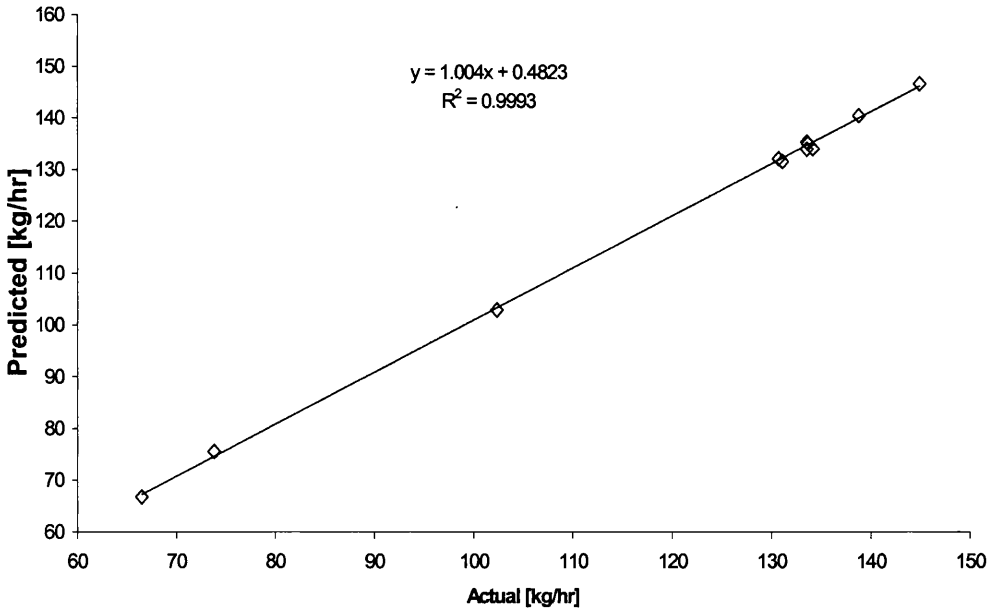


Figure 30 Graph of predicted vs actual throughputs during model confirmation runs

The system was found to be sensitive to the phase of the polymer melt. As long as factor A is kept at a threshold of 110°C or above, the system behaves predictably and with stability. When factor A was set at 90°C and 100°C, the melt exiting the feeding system was observed to be less homogeneous than the melt observed when factor A was set at 130°C - 180°C.

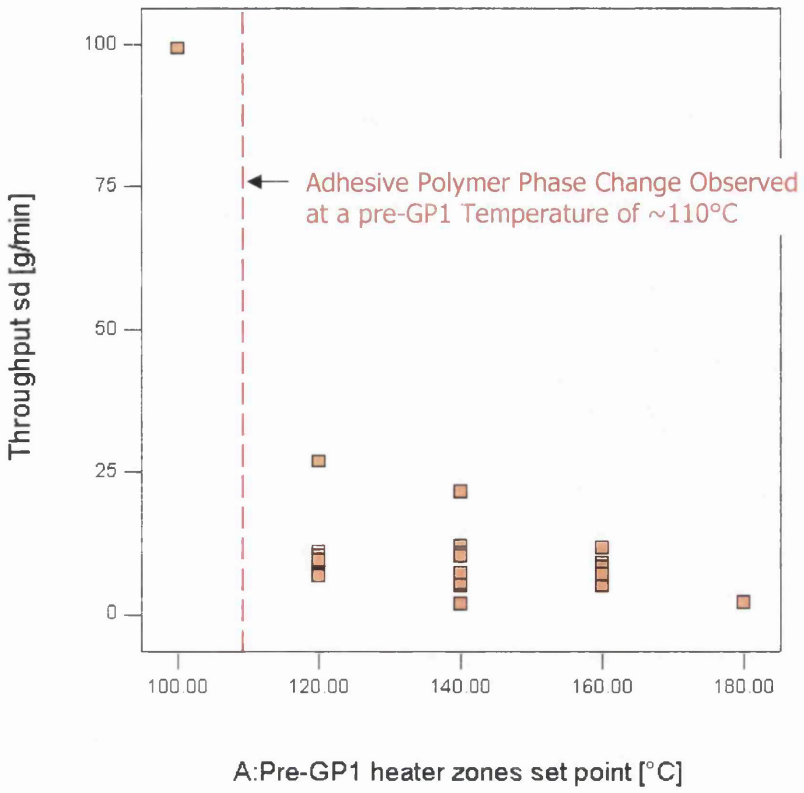


Figure 31 A plot of all of the throughput standard deviation data collected for the DoE

The abnormally high standard deviation comes from the pre-melted polymer experienced when Factor A was set to 100°C (Figure 31). This could indicate that the material had not completely changed phase from its room temperature state, i.e. a gelatinous mass, to a smooth flowing melt. When the melt is more viscous, the gear pump cavities are filled less uniformly giving rise to a lower throughput and a more variable throughput, i.e. at higher temperatures, the melt is more likely to flow and fill the gear pump cavities more uniformly when less viscous.

Maximum throughput is an optimisation goal for this sub-process. The feeding of polymer into the extruder is a rate-limiting step for the full-scale compounding-and-coating process. Using the throughput model, the maximum possible throughput of the system can be established, thus giving a quantifiable upper limit for the whole

process. All the factors are part of the quadratic model equation therefore they all play a part in creating the level of throughput.

Figure 32 graphically shows the effect of factors A and B at optimal levels of factors C and D. It is a plot of factors A & B (the two groups of temperature zones on the X & Y axes respectively). The throughputs are plotted as contours. A GP1 inlet pressure set-point of 19 bar and a maximum pump speed of 65 rpm are held constant. Figure 32 shows that the warmer the settings, the higher the throughput.

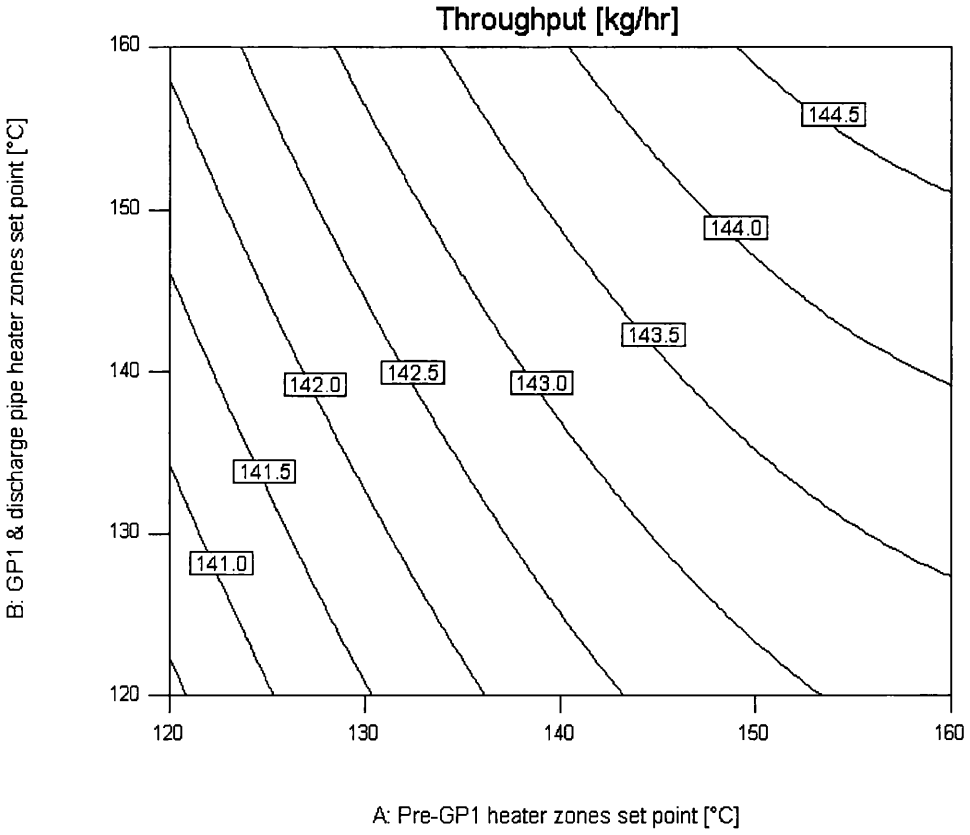


Figure 32 Contour plot of throughput prediction generated from the model derived from the central composite DoE.

The adhesive-polymer-metering system, being part of the compounding system, is required to process adhesive polymer at or above a throughput target. The target throughput is achievable. The DoE analysis predicts a theoretical maximum throughput of nearly 145.0 kg/hr, which is 105% of this target threshold. This part of the process has turned out to be the rate-determining step, i.e. all other equipment is

capable of making the product at a higher rate than this adhesive-polymer-metering system. The implication of this is that in the future it may be necessary to increase the throughput. One possible way of doing this is based on increasing the temperature of the system. This is indicated as an option by extrapolating past the highest throughput, up the gradient of the contours on the response surface plot. The response surface plot is shown in Figure 33. This response surface is the 3D projection of the contour plot shown in Figure 32

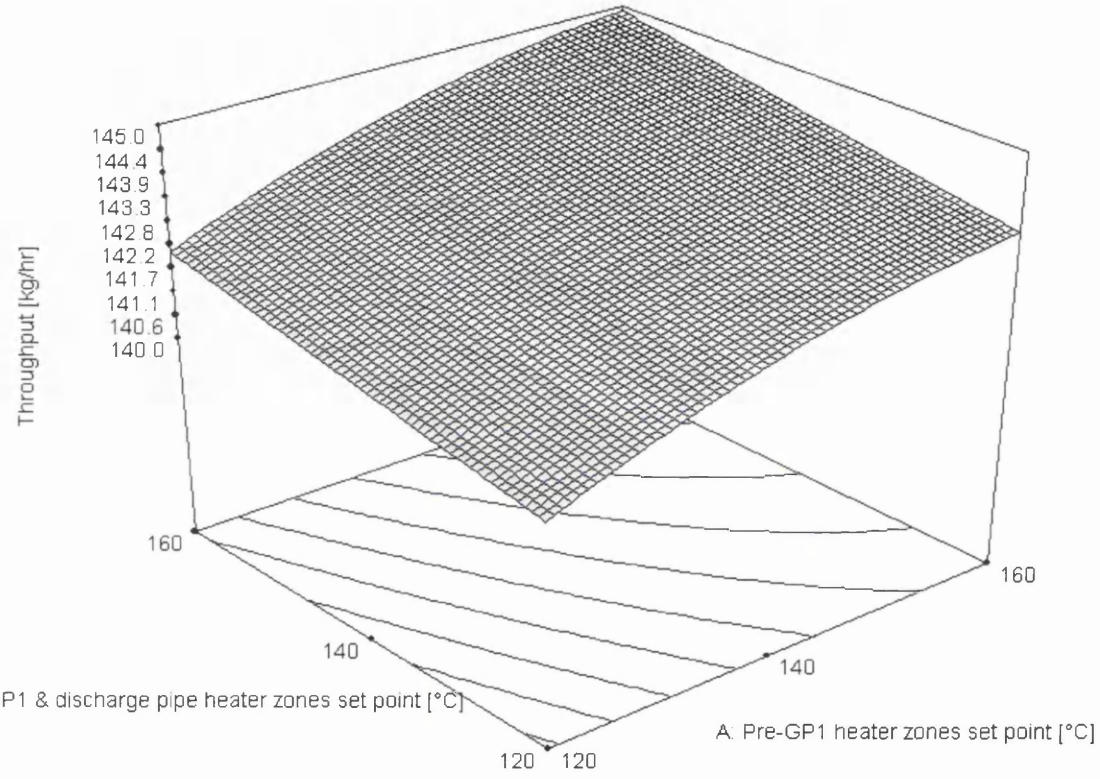


Figure 33 Response Surface throughput prediction generated from the model derived from the central composite DoE.

The contours and gradient suggest there is more throughput to be gained by further increasing the temperatures outside the boundaries of this experiment. However, there is a practical upper limit to the heater set-points, lower than their design maximum. When operating at temperatures above 170°C and the compounding line is stopped, even for a short time, the heaters would over-heat the adhesive polymer. During continuous running, the heaters are set to give an output appropriate to raising the temperature of flowing material. When the flow of polymer suddenly stops, there is

no longer any material taking this heat away. The now excessive heater outputs continue to rapidly raise the temperature of the stationary material and so therefore it overheats. The thermocouples now sense a very high temperature and therefore switch off the heaters in order to lower the temperatures back to the set-points. The machine control logic prevents a restart until all the thermocouples show that the heaters are all controlling within prescribed limits. Because of the machine insulation and low thermal conductivity of the polymer, this takes too long for efficient manufacturing. The implication of this is that a minor cause of stoppage, that may take seconds to rectify, results in 10's of minutes of waiting for the heaters to equilibrate back at normal temperatures. Unnecessary down-time, reduced productivity and therefore increased unit cost results in sub-optimal profit. Any future throughput increase strategy will have to take this in to account.

Analysis of the central composite design for the melt temperature yields a good linear model. The ANOVA that justifies this is shown in Table 12. The linear model equation is shown below and represented graphically in Figure 34.

Melt Temp = 12.79+0.244 * Pre-GP1 heater zones set point [°C] + 0.457 * GP1 & discharge pipe heater zones set point [°C].

As discussed previously, the heater set-points could play an important part in any future throughput increase strategy therefore it is necessary to know whether there is spare melt temperature-raising capacity in the heaters. Melt temperature is independent of GP1 pump speed. This implies that the heaters are powerful enough to raise the temperature of the melt appropriately, whatever the flow rate (pump speed is the largest influence on flow rate).

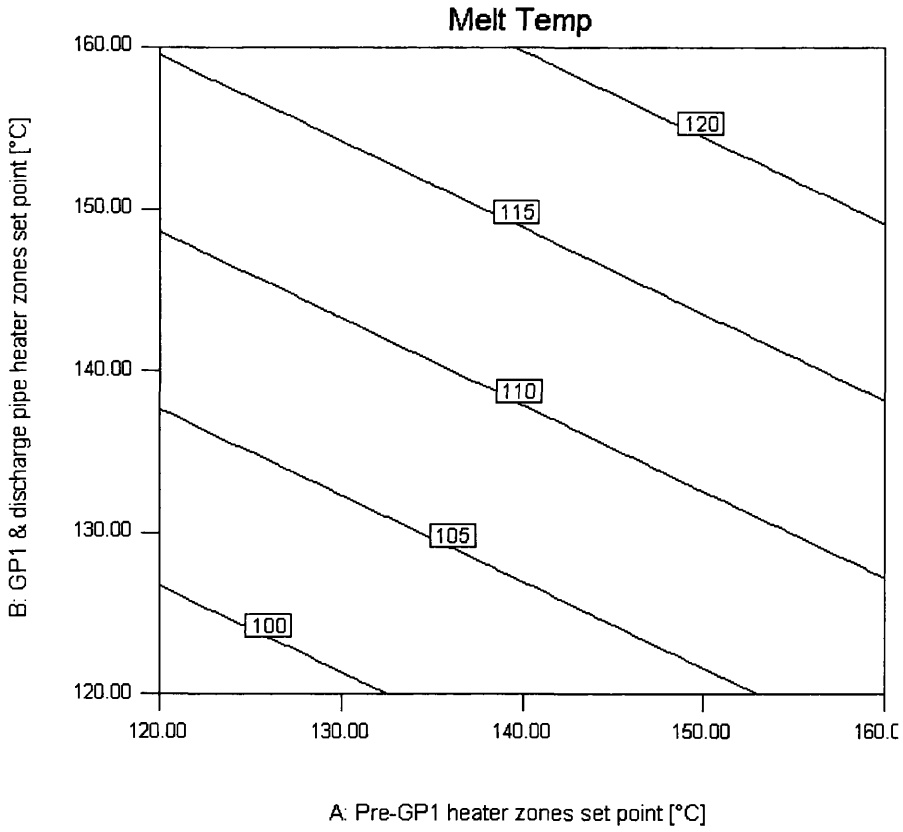


Figure 34 Graphical representation of the linear model for melt temperature.

6.2 Full-scale resin and additive metering system

6.2.1 System Description

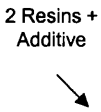


Figure 35 Map of the generic complete process, highlighting resin addition

The resin and additive metering system is the means by which resin and additive are continuously fed into the extruder in a controlled manner. The diagram below details the system.

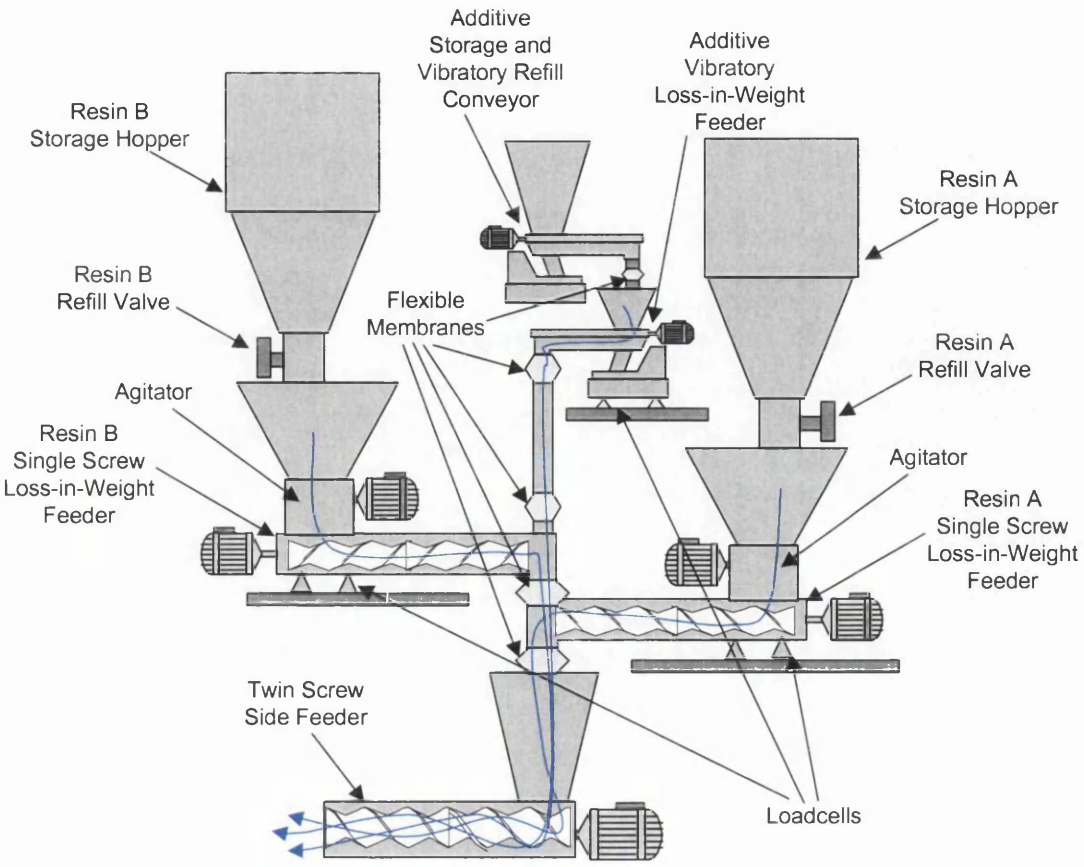


Figure 36 Schematic of the resin and additive feeding system.

The system consists of loss-in-weight (liw) feeders, for feeding and metering resin A, resin B and additive into the compounding twin-screw extruder. Each feeder is controlled from a central microprocessor (Figure 37). Resin A and B feeders convey material using single screws. Screw speed modulation is used to maintain set-point throughput. The additive feeder feeds using a vibratory tray. Vibration amplitude modulation is used to control throughput.

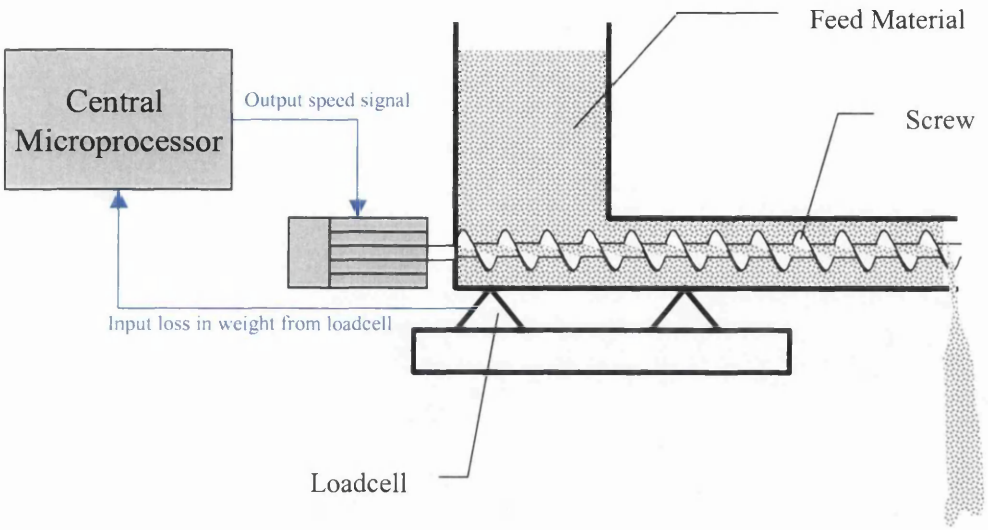


Figure 37 Schematic of a loss-in-weight feeder showing control and feedback

Each liw feeder is mounted on a weigh platform containing loadcells, electronic conditioning units and an amplifier. This system generates a digital weight signal.

Above the feeders, supported on a rigid frame, are hand loaded buffer hoppers, sized to contain approximately four hours of material. Resin A and B buffer hoppers use a slide valve to allow fast filling of the feeders. The additive buffer hopper is fitted with a vibratory tray refill feeder to control refills.

During refills the metering control goes into volumetric mode until the loadcell shock is over, and then returns to gravimetric mode.

Each feeder feeds into a common down-pipe, which in turn feeds the side feeder. The side feeder is starve fed i.e. the speed of the screws is more than fast enough to convey all the material entering them so that there is no backing up. The side feeder pumps the materials into the twin-screw extruder as solids, ranging in size from 10mm crystals down to fines and powders.

The resins are proprietary raw materials supplied to a wide range of companies and industries. The relatively small usage precluded customisation. This also meant that any batch-to-batch variation would have to be absorbed by the process.

The additive is a proprietary 3M product. Limited trials with feeders could be done because of the high cost of the material (about £250 per kg). Most of the feeder vendor trials were done with semolina as an analogous substitute. Although similar in consistency, semolina did not reflect the tendency to form lumps. The feed rate for the additive is towards the lower end of feeder technology capability and therefore dealing with the upset to the control loops that lumps would present needed to be absorbed by the full-scale compounding process.

No preliminary vendor trials with all three feeders in their final configuration were possible as their configuration was unique and there was no mock-up available. From some early investigations it seemed that formulation and therefore feed rate variability had an influence on the finished tape properties.

The SCADA is used to send target throughputs (set-points) to the resin-metering system. The resin-metering system's microprocessor supplies data back to the SCADA for display and analysis. This data has to be correct for the successful compounding of the adhesive. This requires that the feeding system is calibrated, feeds to target and that the material reported to be leaving the system actually enters the extruder (see Figure 38).

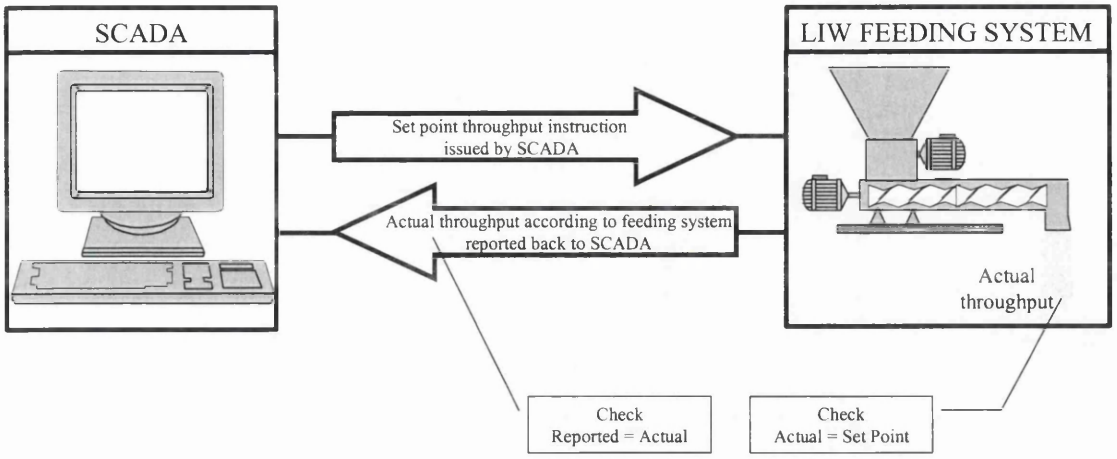


Figure 38 Schematic showing the direction of flow of information between the SCADA and the resin-metering system.

In addition it is necessary to demonstrate that the resin feeders'control system is able to control to target over a prolonged period, taking into account refills and other general operating noise such as particle size changes.

The variability of the resin-metering system also needs to be characterised. Each feeder will contribute some variability to the total compounding throughput, and will therefore add variability to the formulation. As formulation is known to be one of the major influences on product performance, it is essential to characterise the variation inherent in the system, and where possible reduce it. As the resin-metering system throughput set-point is changed during experiments or changes during production, the amount of variation at the differing throughputs needs to be characterised. This tendency for the variability to change as the set-point changes is called variation bias.

The side feeder is designed to pump any resin that enters it into the extruder. If the resin throughput is greater than the throughput of the side feeder, the resins will back-up in the side feeder hopper until contact with a sensor is made. The purpose of the sensor is to alarm such a build up, enabling the operators to take action. If the throughput of the resin-metering system is only just greater than that of the side feeder, it could take a while for the material to back up to the sensor, during which

time the formulation entering the extruder is incorrect. It is necessary to determine when the output from the resin-metering system equals the pumping capacity of the side feeder.

It is necessary to establish:-

- Whether the resin-metering system is calibrated and transferring information to and from the SCADA correctly
- If the resin-metering control system can maintain set-point throughput despite refills and other process noise

6.2.2 Experimental methods

The actual throughput was physically collected in discreet samples, then the sample weights were sequentially plotted on I-MR charts (Individuals and Moving-Range charts).

I-MR charts are a type of statistical control chart that are used for displaying the performance and the spread of measurements from manufacturing systems. They can be used to assess whether the variation generated by a system is predictable or subject to special causes that erode predictability. The more predictable the system the better. A major feature of I-MR charts is that measurments where special causes have acted are highlighted. I-MR charts allow the amount of predictable variation to be defined. They therefore allow the effectiveness of any improvement efforts to be compared. I-MR charts consist of Individuals charts and moving range charts.

Individuals charts are line-plots of data, collected and displayed in sequence, with several additional lines included. The additional lines are the overall mean line, the upper natural process limit (UNPL) and the lower natural process limit (LNPL) lines. The overall mean line is simply the mean of the data. The natural process limits are

estimates of ± 3 σ (i.e. the limits between which 99.7% of the data can reasonably be expected to be distributed, assuming that the data comes from a normal population). This is shown in the example below.

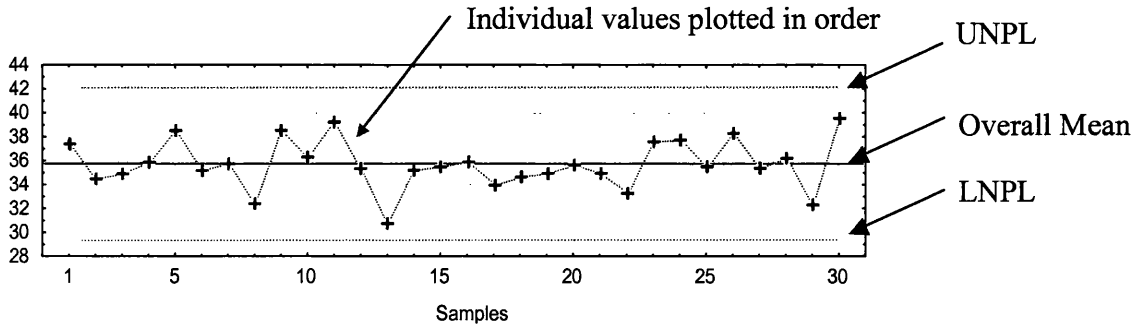


Figure 39 Sample I-MR chart

The moving range chart is a similar plot except the individuals are replaced by the difference between the two consecutive individual readings i.e. the moving range.

The spread of data within the NPL's is also important and can be used to assess the normality of the data. Deviation from a random normal pattern of data suggests a strong likelihood that the process is subject to the influences of a special cause. Eight tests have been developed to highlight special causes. These are shown in Figure 27. They rely on the improbability of getting certain patterns of data from a properly random normal distribution.

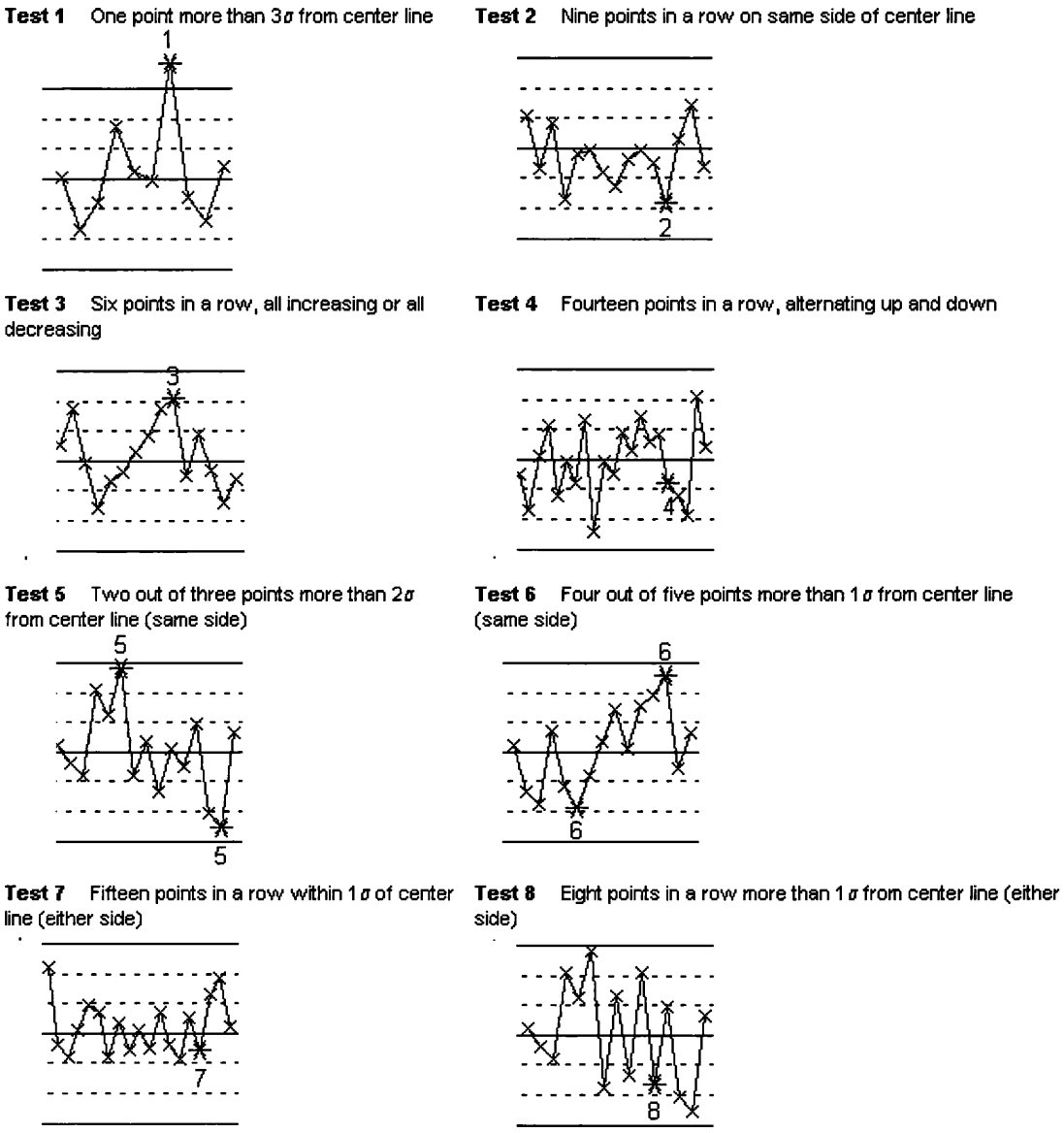


Figure 40 The eight tests for special causes[16]

An example of one of the tests is Test 3 "6 points in a row all increasing or all decreasing". If 6 points in a row are steadily increasing or decreasing then this signals a drift in the process average. Often, such drift can be the result of wear, deteriorating maintenance, improvement in skill, etc.

The SCADA issues an instruction to the resin-metering system to feed at a certain throughput. To test that the instruction has been correctly issued and executed, I-MR charts were used. The rationale for their use follows.

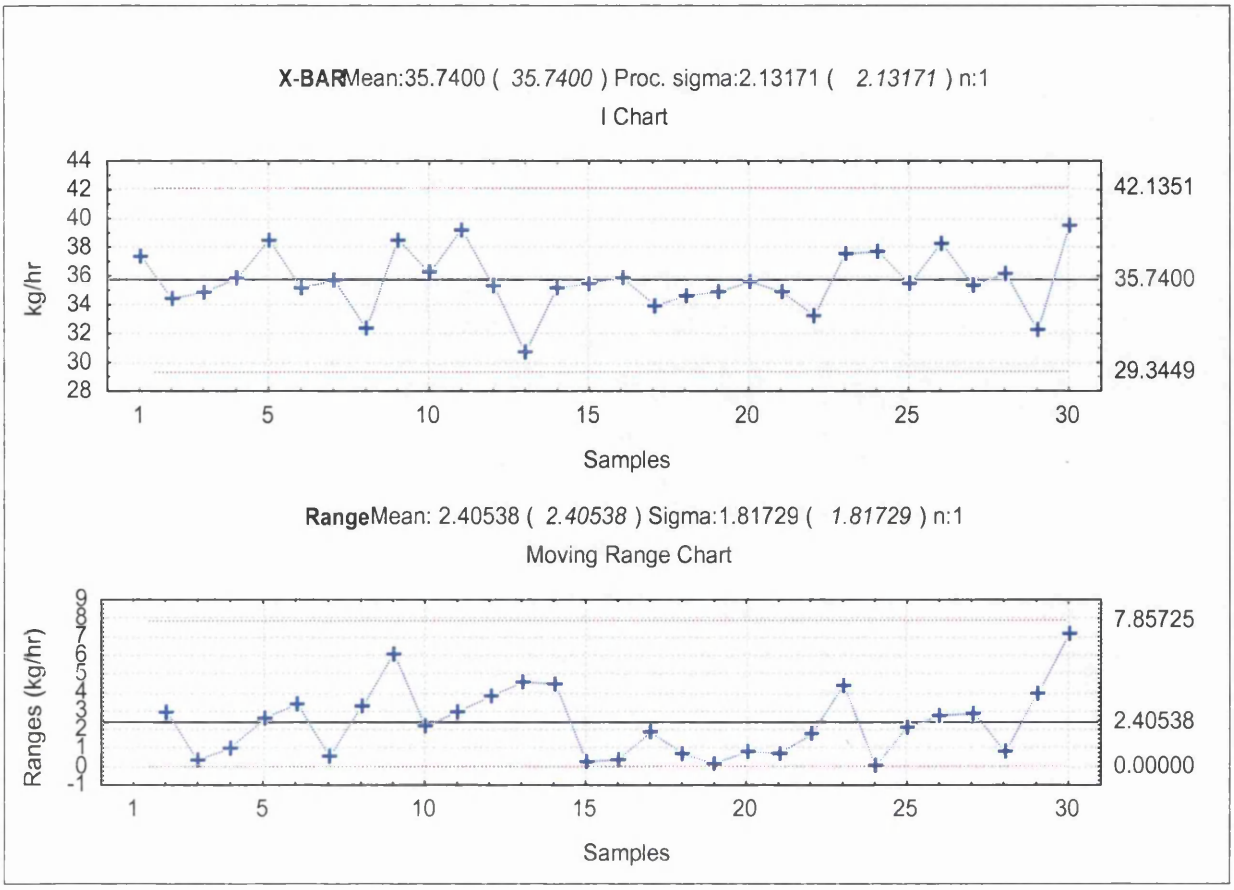
If the resin-metering system throughput data is "in control", i.e. no special causes are detected by an I-MR chart, then the overall mean of the data can legitimately be compared with the set-point throughput, i.e. the data was collected from a steady state system and therefore the mean represents the system accurately. Difference between the overall mean and the set-point can be evaluated for significance by using a two-sided t-test. If the t-test shows a statistically significant difference then this would indicate a discrepancy between set-point and actual. This would mean that either the SCADA had not issued the instruction properly or the feeding system had not carried it out. Possible causes of this type of failure that would then need to be investigated are calibration errors or data communication errors. If there were no statistically significant difference then this would validate the flow of information from SCADA to the resin-metering system.

I-MR charts can be used to problem solve any abnormal resin-metering behaviour problems using the tests for special causes. If the resin-metering system throughput data is not "in control", i.e. special causes are detected by an I-MR chart, then the overall mean of the data can not legitimately be compared with the set-point throughput, i.e. the data was collected from a dynamically changing-state system and therefore the mean does not represent the system accurately. In this case I-MR charts, because the data is plotted sequentially, can be used to pinpoint the nature of the lack of stability and therefore possibly the cause. Inferences can be drawn from the type of special cause. For example a test 1 special cause, i.e. one point more than 3 sigma from the mean, may represent a shock to the loadcell due to a knock, momentary vibration from another operation or an unusually large particle of resin being fed.

The resin-metering system sends information to the SCADA. To establish if the set-point throughput equals the reported throughput, a three factor DoE was run. The factors were a] ratio of Resin A to adhesive polymer, b] ratio of resin B to adhesive polymer and c] ratio of additive to adhesive polymer. The levels were chosen to reflect the likely extremes of future operation. Some of the responses analysed included the recorded throughput and recorded throughput standard deviations, logged by the SCADA. Because of the duration of the experiment, the measurements included refills and other process noise, such as having to meter a range in particle sizes. If the reported values are the same as the actual values then the derived model will show no other influence apart from the feeders' own set-points. This experiment was also designed to look for variation bias. If there is variation bias, i.e. the amount of variation a feeder generates is dependent on the throughput of that feeder, a relationship between set-point throughput and standard deviation will exist. Lack of a model will imply lack of variation bias over the ranges covered by the experiment.

6.2.3 Resin-metering System Characterisation

The I-MR results for feeder A are shown in Figure 41. The I-MR control chart shows statistical control i.e. no special causes have been detected by the eight tests. Because feeder A is in statistical control it is valid to carry out the t-test. Using a rejection level of 0.05, the t-test shows that the average of the individual measurements equals the set-point (hypothesised mean). This is confirmation that the manually measured throughput equals the set-point throughput for feeder A. As there are no special causes, refills or any other momentary process noise do not influence the metering of this system.

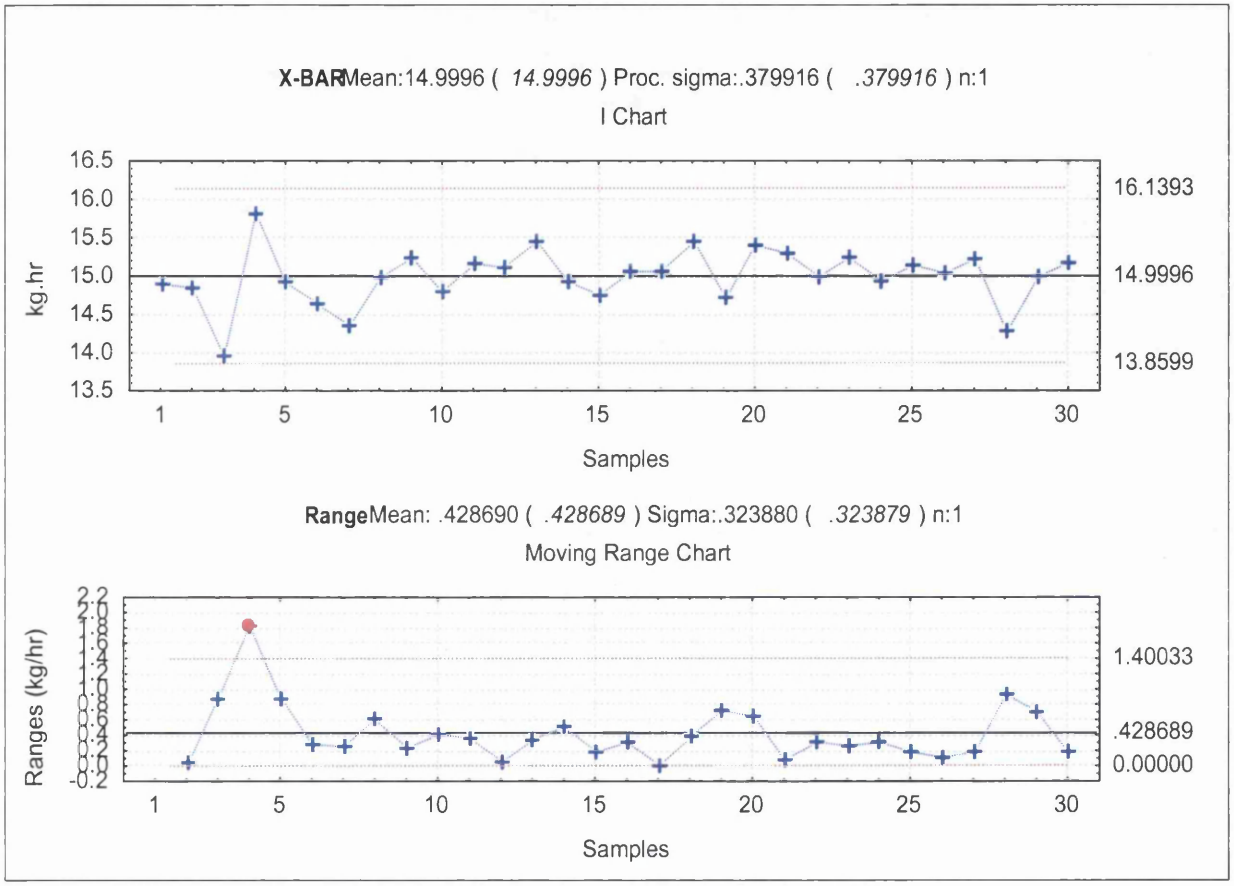


N	Hypothesised mean (Set Point)	Mean	Std. Dev.	Std. Err.	t	df	2 sided P val
30	35.08	35.74	2.0495	0.3742	1.7639	29	0.088290511

Figure 41 I-MR Control charts for feeder A and "two sided test alternative" t-test for feeder A

The I-MR results for feeder B are shown in Figure 42. The I-MR control chart shows that statistical process control is demonstrated by the individuals, however the range chart shows 1 point outside of the upper control limit (Test 1 of the eight special cause tests is positive). This special cause is probably due to timing error i.e. one sample was removed from the stream prematurely, consequently the next sample had the previous samples shortfall of material added to it. This is backed up by the individuals chart showing one very low result (point 3) followed by a very high result (point4). If this is the case then the overall average will not be affected by the special cause. The appended t-test results show that the average equals the set-point (hypothesised

mean). This is confirmation that the manually measured throughput equals the set-point for feeder B. With the special cause data removed, the increased P value of the t-test shows increased reason to accept the null hypothesis. As there are no special causes, refills, or any other momentary process noise do not influence the metering of this system.

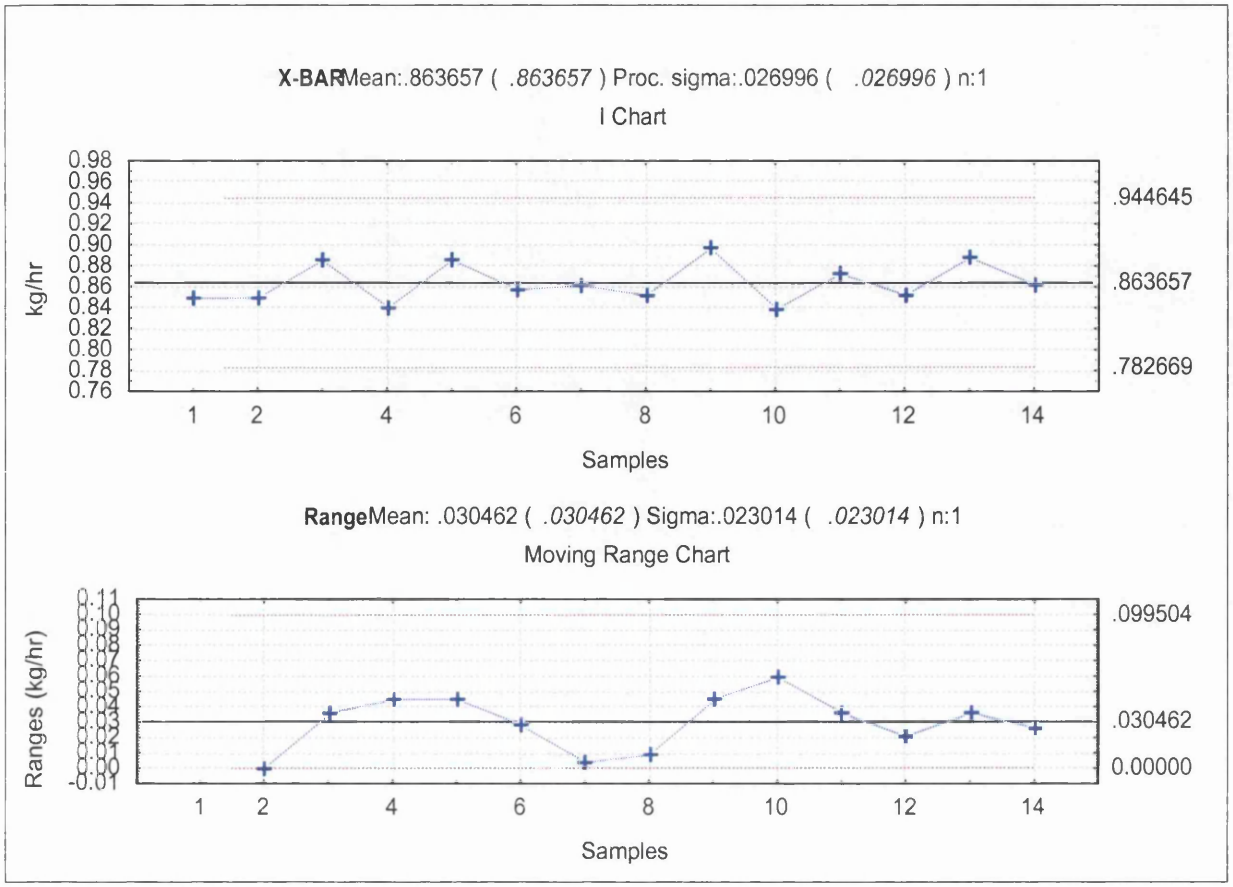


N	Hypothesised mean (Set Point)	Mean	Std. Dev.	Std. Err.	t	df	2 sided P val
30	15.04	15	0.3672	0.0670	-0.6026	29	0.5515

Figure 42 I-MR Control charts for feeder B and "two sided test alternative" t-test for feeder B

The I-MR results for the additive feeder are shown in Figure 43. The I-MR control chart shows statistical control and therefore it is valid to carry out the t-test. The t-test shows that the average of the manual measurements equals the set-point (hypothesised mean). This is confirmation that the manually measured throughput equals the set-point throughput for the additive feeder. As there are no special causes,

refills, or any other momentary process noise do not influence the metering of this system.

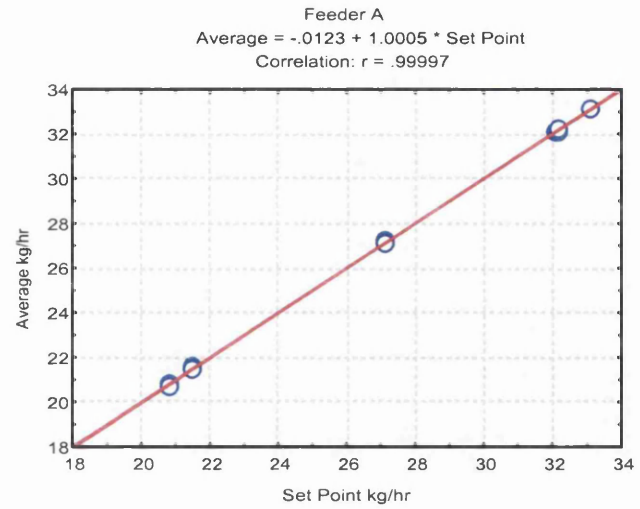


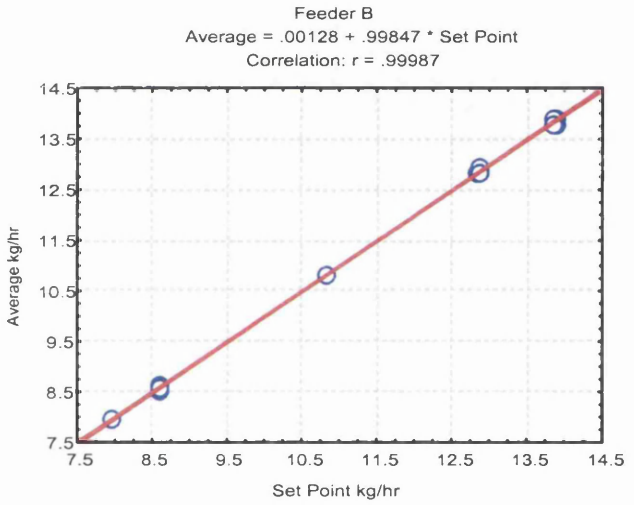
N	Hypothesised mean (Set Point)	Mean	Std. Dev.	Std. Err.	t	df	2 sided P val
14	0.86	0.8637	0.0192	0.0051	0.7145	13	0.4875

Figure 43 I-MR Control charts for the additive feeder and "two-sided test alternative" t-test for the additive feeder

The set-point throughput equals the throughput reported by SCADA.

Data from equipment commissioning trials showed that for each feeder the only factor that influences a feeder is its own set-point. There is a good correlation between the set-point throughput and mean throughput reported to SCADA and the relationship is 1:1. See the correlation plots in Figure 44. The gradients are close to 1, the intercepts are close to 0 and the correlation coefficients are close to 1. This means that the set-point throughputs equal the throughputs reported by SCADA for all three feeders.





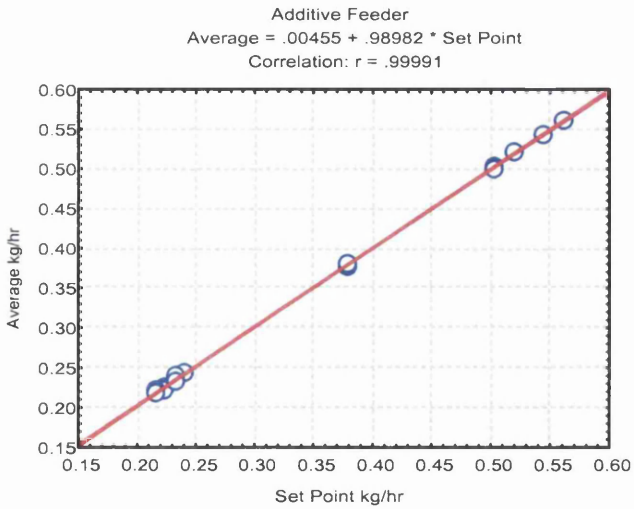


Figure 44 Correlation between feeder set-point throughputs and throughputs reported to SCADA for feeders A and,B and the additive feeder.

There is no variation bias over a reasonable range of throughput set-points. No model could be derived from the DoE data. This suggests that throughput standard deviation is unaffected by the factors chosen for the experiment and can not be optimised using these factors. Therefore the throughput set-point does not affect variation and therefore there is no bias due to variation.

6.3 UV curing

The UV curing system was scaled from a laboratory-scale bench-top model to the final design for the full-scale coating process. The modelling in this section was to prove the assumptions in the extrapolation from lab-scale to the full-scale design. These assumptions formed the basis of a specification for the full-scale system. The specification included a degree of over capacity in case the assumptions were wrong. Therefore once the system was installed, the system was not run at full power to avoid excessive web temperatures, heat being a by-product from the lamps. This required further experimentation with parameters not available on the lab-scale system.

6.3.1 UV Curing System Description

The coated compounded adhesive requires UV radiation to initiate a chemical modification that leads to the desired tape properties. To achieve this, the full-scale process has two UV curing chambers immediately after each coating station. Each chamber has a 2x2 array of UV lamps. The lamps are a standard width, one lamp is too short to span the web, two lamps butted up to each other are long enough to span the web with some overhang. This overhang represents wasted energy as there is no substrate under it. To overcome this waste, each row of lamps is angled such that there is no overhang and therefore no wasted radiation (Figure 45).

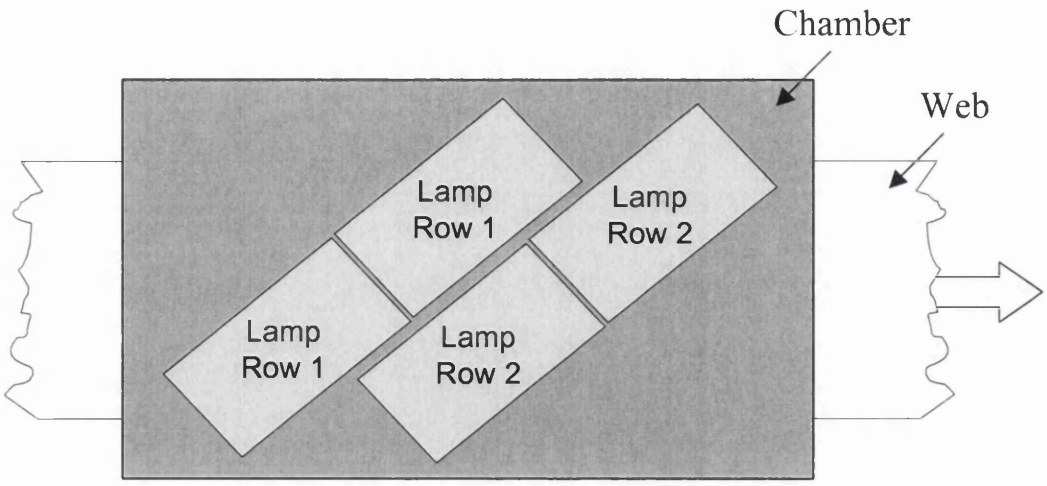


Figure 45 Top view of a curing chamber.

Each lamp is controlled by its own power supply unit and is capable of about 240W/cm (6kW total) at maximum power. The power is adjustable between 25% and maximum, the higher the power setting, the greater the intensity of the lamps and the greater the radiated heat that can distort the web. The set-point percentage power for production or experimentation is sent from the SCADA (Figure 46). Each lamp in a row (cross-web pair) is matched electronically, and will irradiate at the same percentage power, however the percentage power of each row can be set independently. There is no on-line dose sensor even though the key influence on the cure of the product is dose.

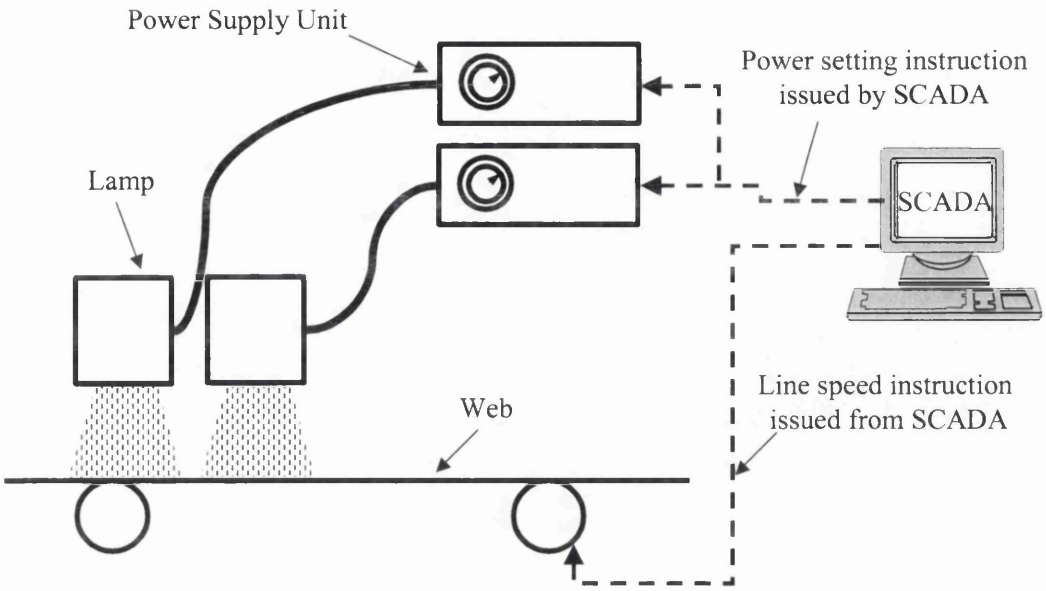


Figure 46 Schematic of the relationship between the SCADA and the full-scale UV system.

The UV curing system fitted to the full-scale coating process is also available in a small laboratory-scale version, albeit with several major differences. The most significant differences are the power rating of the lamp, i.e. 120W/cm compared with the full-scale system that is rated at 240W/cm, and the lamp is not angled, i.e. it is perpendicular to the web.

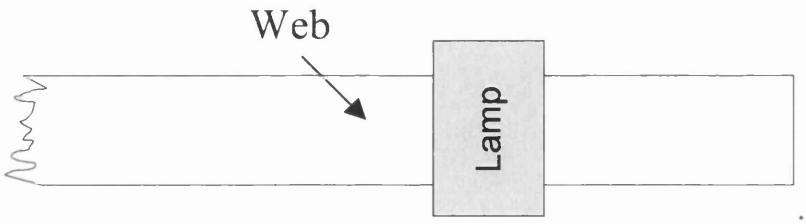


Figure 47 Laboratory-scale UV system from above

6.3.2 UV Radiation Modelling

Preliminary experiments, run in a laboratory, to aid the design of the final UV curing system are described in this section followed by a designed experiment on the full-scale system.

The laboratory trials were necessary to produce a specification that ensured that the purchased equipment would be capable of achieving the target UV dose of 800mJ/cm^2. The issues this addressed were, could a useful model of the UV curing system be derived from trials run on laboratory equipment that would enable a specification for full-scale equipment to be written with a high level of confidence. If so, what is the most economic configuration of lamp types, numbers of rows of lamps and orientation of lamps that would provide an adequate curing process window. Design confirmation trials were then required on the full-scale equipment.

Once the design confirmation trials had confirmed that the full-scale system was able to meet the requirements demanded of it, a further investigation was necessary. The objective of these experiments was to gain the control of the process variables and understanding their influence on UV dose as there is no on-line dose sensor. In this investigation dose is a response but in later process optimisation experiments, dose is a factor. For an optimised system it should be possible for the SCADA to set the correct lamp power to give the correct dose at the set line speed, therefore the dose, power, line speed relationship is required. Excessive dose leads to excessive heat from the lamps. This heat may distort the web as it contains uPVC. UPVC has a tendency to shrink at temperatures in excess of 70°C. Shrinkage would cause web curl and other web-handling issues. This investigation, a DoE, was required to include a process variable not available on the laboratory-scale system, in the UV curing model. This additional variable is lamp %power adjustment. The laboratory-scale system did not have the ability to alter the %power. It was reasonable to assume that the relationship between %lamp power and dose may not linear. Hence the requirement for a DoE that could model curvature.

6.3.3 Derivation of a model for specifying full-scale process equipment.

The UV dose was measured using an UV Power Puck form EIT Instruments. The device is a multiple wavelength power meter that displays total dosage (0-250joules/cm^2) and peak intensity (0-5W/ cm^2). It does so over four UV ranges measured simultaneously: UVA (320-390nm), UVB(280-320nm), UVC(250-260nm), UVV(395-445nm). It can measure peak intensities in the range of 5mw/cm^2 to 5W/ cm^2, outputting dose and intensity to a LCD display.

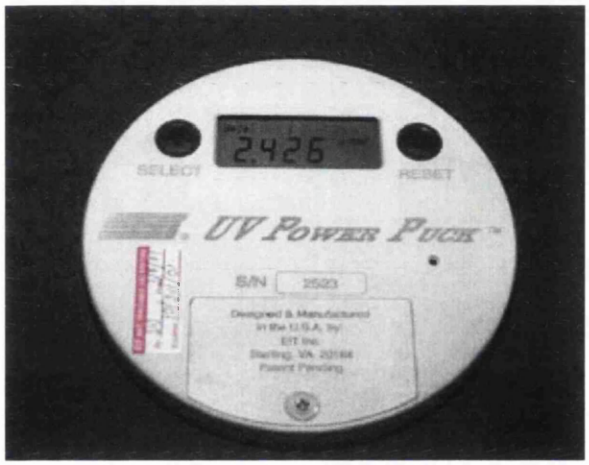


Figure 48 UV Power Puck

In this investigation UV dose data was generated using a laboratory UV curing system shown in Figure 49. The UV measurement device was placed on the conveyor travelling at varied speeds, exposing the Puck to the lamp. The UV radiation impinges on the Puck's UV-sensitive optical detector and automatically measures and records UV energy levels (as a function of time) which equals the dose experienced by the measurement device, and therefore the web. The UV detector is designed to measure pre-set bands of wavelengths within the UV spectrum. The optimum curing of the tape occurs at around 350nm, therefore the pre-set band of 320-390nm, or UVA band, is the most relevant instrument setting to log data.

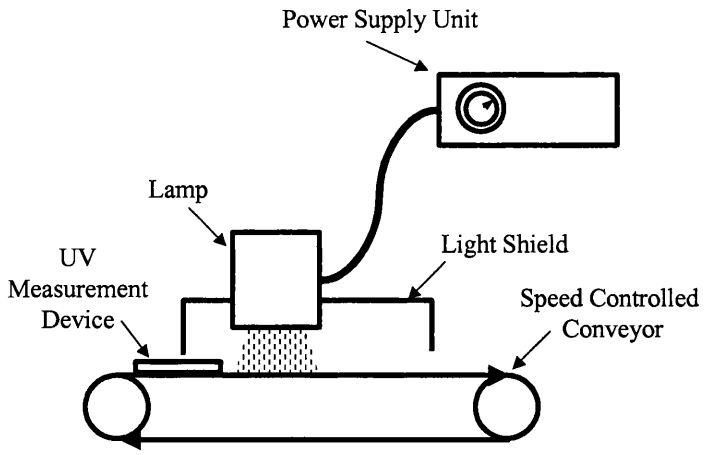


Figure 49 Schematic of the laboratory UV curing system.

Conveyor speed was the only factor in this experiment; all other variables were fixed, such as the lamp focus (fixed at optimum). To adapt the lab model to the full-scale UV curing system, certain assumptions were made:

1. The dose is doubled by doubling the number of lamps i.e. the total dose from an array is the sum of the doses produced by each row of lamps.
2. The dose is doubled by doubling the power i.e. the 240W/cm system produces twice the dose of the laboratory 120W/cm system.

3. The dose can be increased by angling the lamps. The increase = 1/sin θ°. See Figure 50.

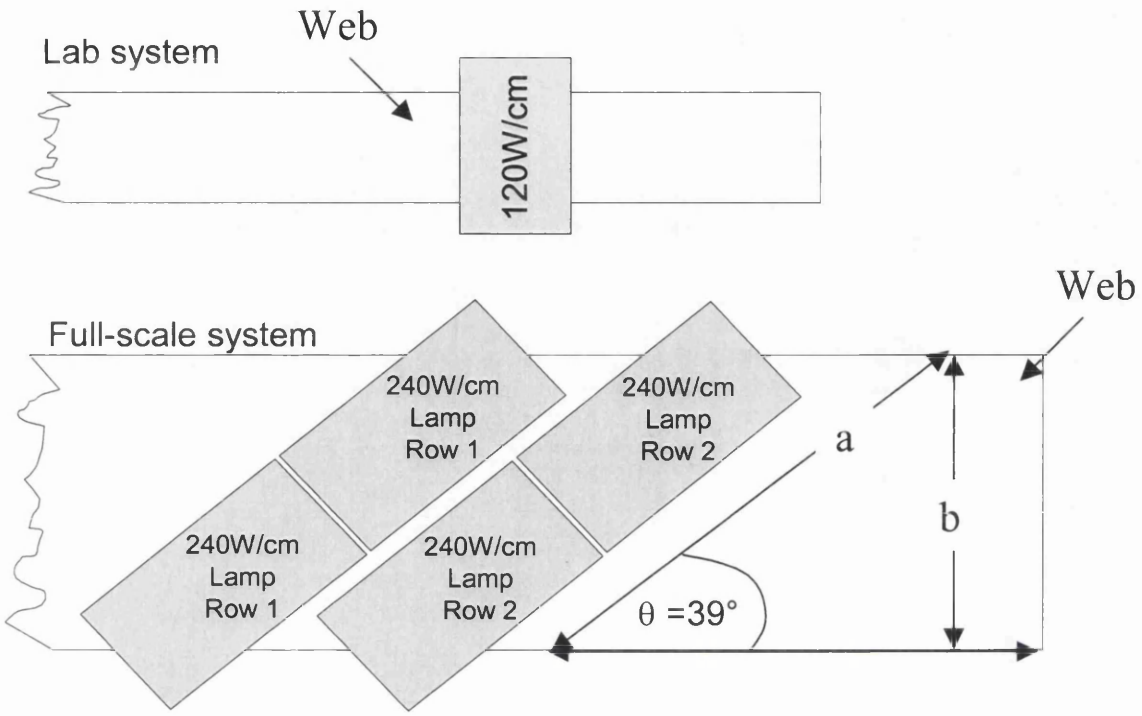


Figure 50 Schematic showing the laboratory UV curing system and the full-scale UV curing system.

The results of the single factor experiment are shown below

Belt Speed	Measured UVA Dose [J/cm²]	1/Dose [cm² /J]
5	1.65	0.61
10	0.82	1.22
15	0.56	1.79
20	0.44	2.27
25	0.35	2.86
30	0.29	3.45
35	0.24	4.17
40	0.22	4.55
45	0.18	5.56
50	0.18	5.56
55	0.17	5.88
60	0.14	7.14
65	0.13	7.69

Table 13 Results of measuring UVA dose at various belt speeds.

There is a strong relationship between the conveyor speed and measured dose. It is an inverse relationship i.e. at zero conveyor speed there is infinite dose and at infinite speed there is zero dose. The measured dose data was inversed and modelled using a linear equation, derived from linear regression. The ANOVA for which is appended below.

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	60.419	1	60.419	1314.801	< 0.0001
Residual	0.505	11	0.046		
Cor Total	60.925	12			

R-Squared	0.992
Adj R-Squared	0.991
Pred R-Squared	0.988
Adeq Precision	82.230

Table 14 ANOVA for the single factor model of inverse dose

The F Value associated with the model is >1314. The Prob > F value is <0.0001, therefore the model effect is large. This value implies that the chances of getting an F value this large if the result was due to noise is less than 0.01%. Therefore the result is almost certainly due to altering the factors.

The adjusted R squared value is >0.99, therefore the model explains the majority of variation observed during the experiment.

The graph in Figure 51 shows the individual data points and the UV dose model as the line. The R^2 index, being close to 1 suggests that the model explains the overwhelming majority of the variation. The model is described by equation 1.

*Equation 1. Dose [J/cm²] = 1/(0.1152*Conveyor Speed+0.023)*
Where conveyor speed is measured in m/min.

The strong relationship confirms current knowledge that the dose is inversely proportional to speed.

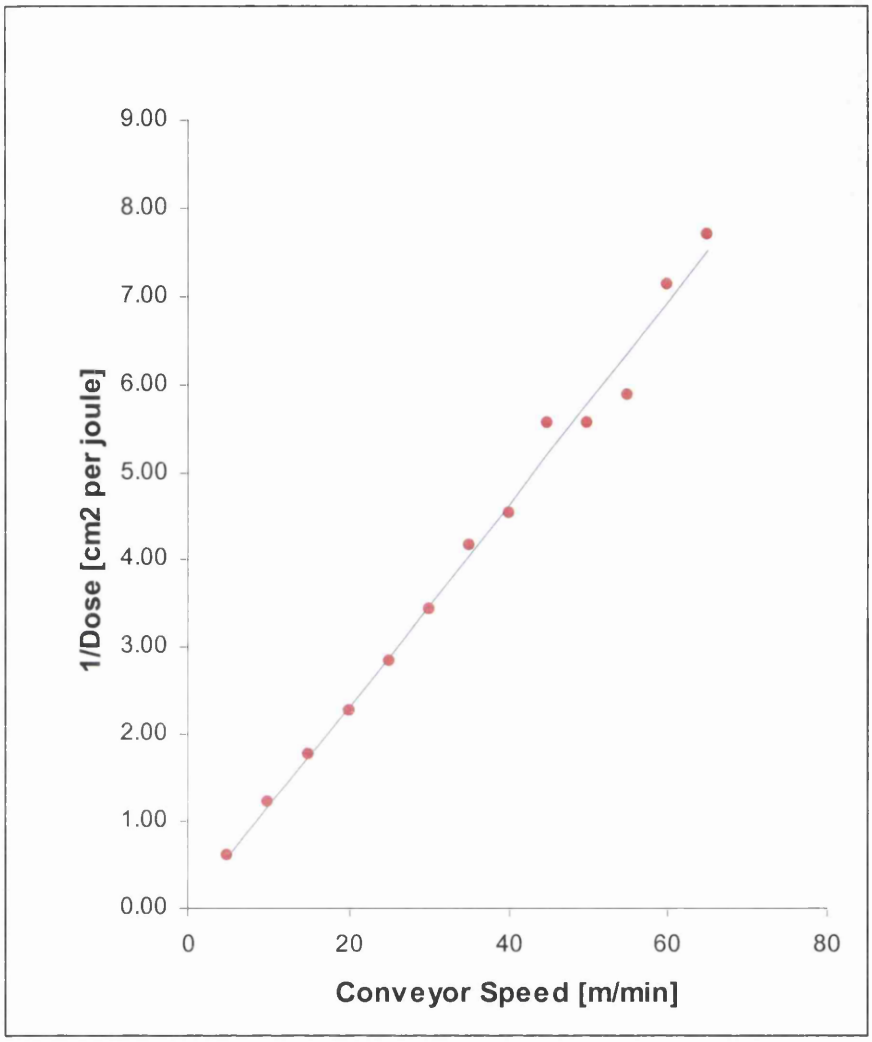


Figure 51 Relationship between line speed and reciprocal dose using the laboratory-scale 120W/cm system.

Having modelled the relationship between conveyor speed and dose, the scaling assumptions can be applied to give a theoretical model that can be used to specify the full-scale system. Applying the assumptions one by one alters the model Equation 1 as follows:-

Assumption that doubling the power of the system doubles the dose

Equation 2:

Dose [J/cm²] = (1/(0.1152*Conveyor Speed+0.023))***(Lamp Power Rating/120)**

Where the lamp power rating is measured in W/cm .

Assumption that angling the lamps alters the dose in proportion to 1/sin θ°

Equation 3:

Dose [J/cm²] = (1/(0.1152*Conveyor Speed+0.023))*(Lamp Power Rating/120) ***(1/sin θ)**

Where θ is the angle described in Figure 50.

Assumption that increasing the number of lamps alters the dose in proportion to the number of lamps

Equation 4

Dose [J/cm²] = (1/(0.1152*Conveyor Speed+0.023))*(Lamp Power Rating/120) *(1/sin θ) *** Number of Lamps**

Equation 4 can be used to calculate the optimum chamber configuration, satisfying the requirements that each chamber must produce at least 0.8J/cm^2 at 50 m/min. The best solution is 2 rows of 240W/cm lamps angled at 39° per chamber i.e. option 8 in Table 15. This arrangement has no wasted overhang, the smallest number of lamps (therefore cost), whilst having a 38% safety margin of dose to allow for any inaccuracies in the model. This is the option that was chosen and incorporated in to the full-scale process.

Option	Power rating [W/cm]	Lamp unit width [mm]	Lamp Orientation	Min num. lamps required per row	a [mm]	Web Width b [mm]	Wasted overhang [mm]	Angle theta [°]	Min num. of num rows	Max Theoretical Dose [J/cm²]	Total Num of lamps
1	120	152.4	Perpendicular	3	320	320	137.2	90.0	5	0.86	15
2	120	152.4	Angled	3	457.2	320	0	44.4	4	0.99	12
3	120	254	Perpendicular	2	320	320	188	90.0	5	0.86	10
4	120	254	Angled	2	508	320	0	39.0	3	0.82	6
5	240	152.4	Perpendicular	3	320	320	137.2	90.0	3	1.04	9
6	240	152.4	Angled	3	457.2	320	0	44.4	2	0.99	6
7	240	254	Perpendicular	2	320	320	188	90.0	3	1.04	6
8	240	254	Angled	2	508	320	0	39.0	2	1.10	4

Table 15 Lamp types available, and their configurations that allow a web 320mm wide, travelling at a line speed of 50m/min, to receive at least 0.8J/cm² .

After the full-scale UV curing system was installed, the laboratory model was checked by making several measurements using a UVImap measurement device, sensitised to the UVA band (320-390nm), at various line speeds under lamps in different configurations. The UVImap is a similar device to the UV Power Puck described in section 6.3.2 except it also has a thermocouple that can measure the temperature the device experiences during a data collection cycle. The data is not displayed on a digital display Instead the device digitises and stores in random access memory up to 4000 UV and 2000 temperature measurements taken during the data collection cycle. Upon completion of a cycle, the instrument is connected to a printer where the total UV dose, peak value of UV intensity, average UV intensity, UV sample rate, peak value of temperature, and the number of samples points is printed along with an intensity verses time plot and a temperature verses time plot.

Each assumption was tested and the results plotted in Figure 52. The blue line represents the model's predictions. The red circles represent the data that was used to generate the original model. The various lamp configurations are annotated.

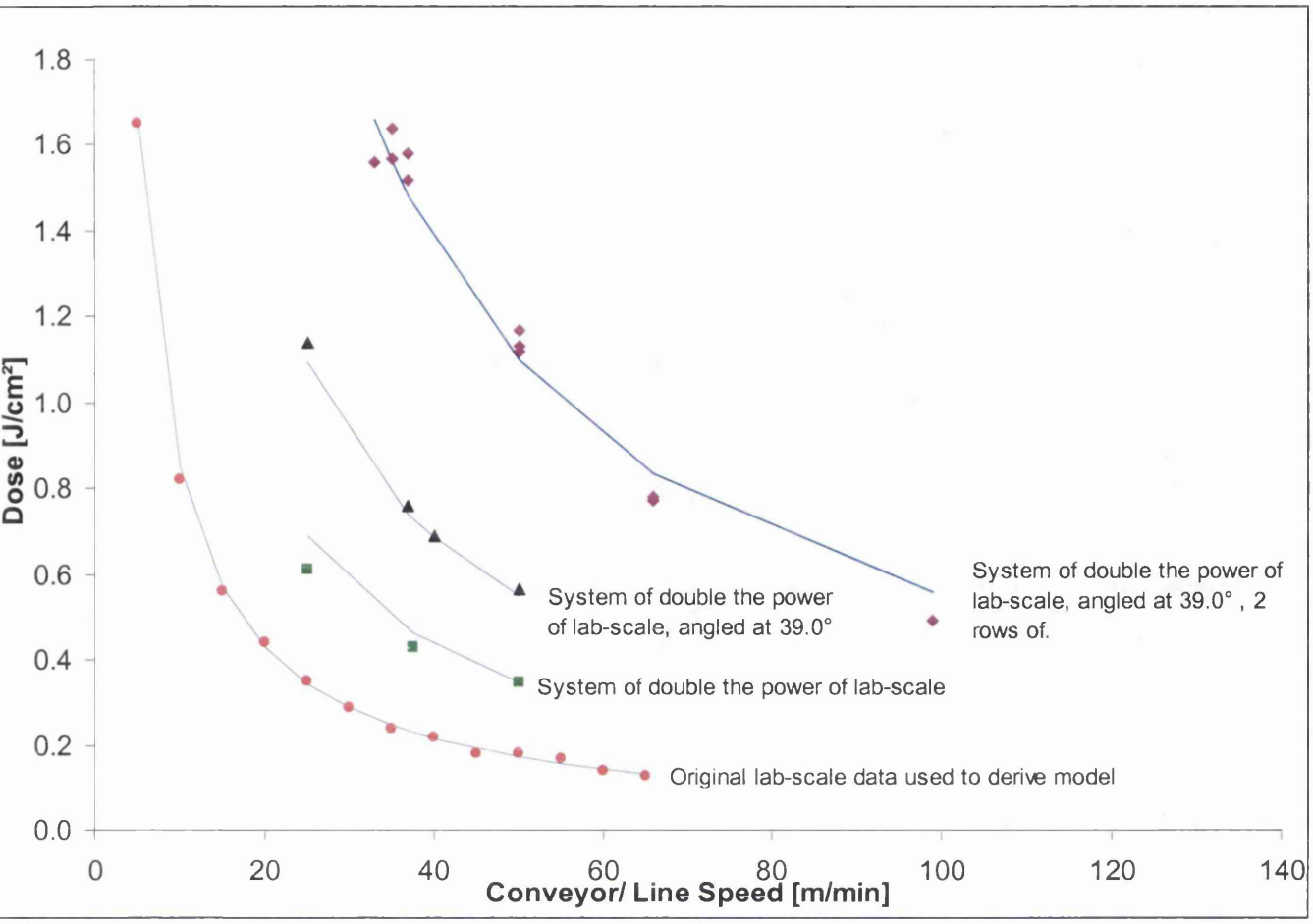


Figure 52 Plot of predictions of different lamp types in different configurations

Figure 53 is a correlation plot of the actual data and the predicted data. The fact that the gradient is close to 1 and the intercept is close to zero implies that there is overall agreement between the two sets of data. The R^2 value, being close to 1, implies that this model is precise and therefore the original assumptions included in the model are sound, making the model trustworthy.

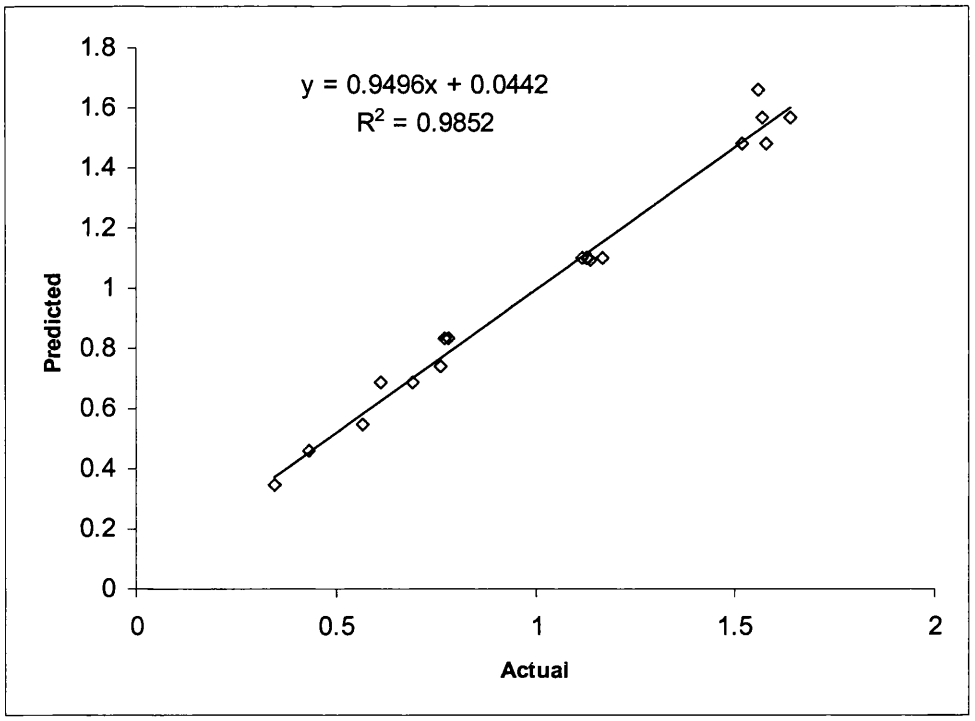


Figure 53. Correlation plot of the model prediction and the actual values measured at various line speeds with different configurations of lamps.

6.3.4 Derivation of a model involving the % lamp power

Table 15 describes how the predicted dose of the full-scale system, using validated models, is 138% of the requirement of 800mJ/cm². This excessive dose may cause web-handling problems as the excess heat generated may cause the uPVC in the web to curl. Once the full-scale UV curing system became operational, a new process variable, % lamp power, became available to adjust lamp power. The line speed / dose relationship is not linear, nor is the expected relationship between % power and dose. Therefore a DoE that could model curvature, potentially with quadratic equations, was run. This section describes the experiment, the findings and the implications.

6.3.4.1 Experiment and findings

The experiment was a central composite DoE. The factors were web speed and % lamp power set-point. The responses were UVA dose and web temperature, measured

by placing the measurement device on the moving web, passing it under one row of lamps. The measurement device has a thermocouple attached to collect temperature data. The experimental layout and results are shown below.

Std	Run	Type	A:Web speed [m/min]	B:%Lamp power	Dose J/cm²	Web temp [°C]
1	9	Factorial	25	50	0.2715	74.4
2	1	Factorial	50	50	0.141	60.8
3	3	Factorial	25	100	1.14	116.8
4	6	Factorial	50	100	0.566	121.6
5	2	axial	25	75	0.78	88
6	8	axial	50	75	0.386	79.2
7	4	axial	37	50	0.1895	69.6
8	5	axial	37	100	0.76	122.4
9	7	Centre	37	75	0.527	90.4

Table 16 Central composite DoE and results.

6.3.4.1.1 Dose

The ANOVA for the dose response shows that a good model can be derived

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	0.827175	3	0.275725132	83.85748	0.0001
A	0.199395	1	0.199395497	60.64302	0.0006
B	0.57346	1	0.57345977	174.4088	< 0.0001
AB	0.048697	1	0.048697233	14.8105	0.0120
Residual	0.01644	5	0.003288021		
Cor Total	0.843616	8			

R-Squared	0.980512
Adj R-Squared	0.96882
Pred R-Squared	0.952287
Adeq Precision	25.71161

Table 17 ANOVA for dose

The F Value associated with the model is >83. The Prob > F value is 0.0001, therefore the model effect is large. This value implies that the chances of getting an F value this large if the result was due to noise is 0.01%. Therefore the result is almost certainly due to altering the factors.

The adjusted R squared value is >0.96, therefore the model explains the majority of variation observed during the experiment. The model is described by the following equation:-

Equation 5

Dose [J/cm²] = -0.847 +0.0119 * Web speed+0.0256* %Lamp power-3.53E-004*Web speed* %Lamp power

Where web speed is in m/min.

A graphical representation of this equation is shown in Figure 54. The contours are dose and the red circles are the conditions run.

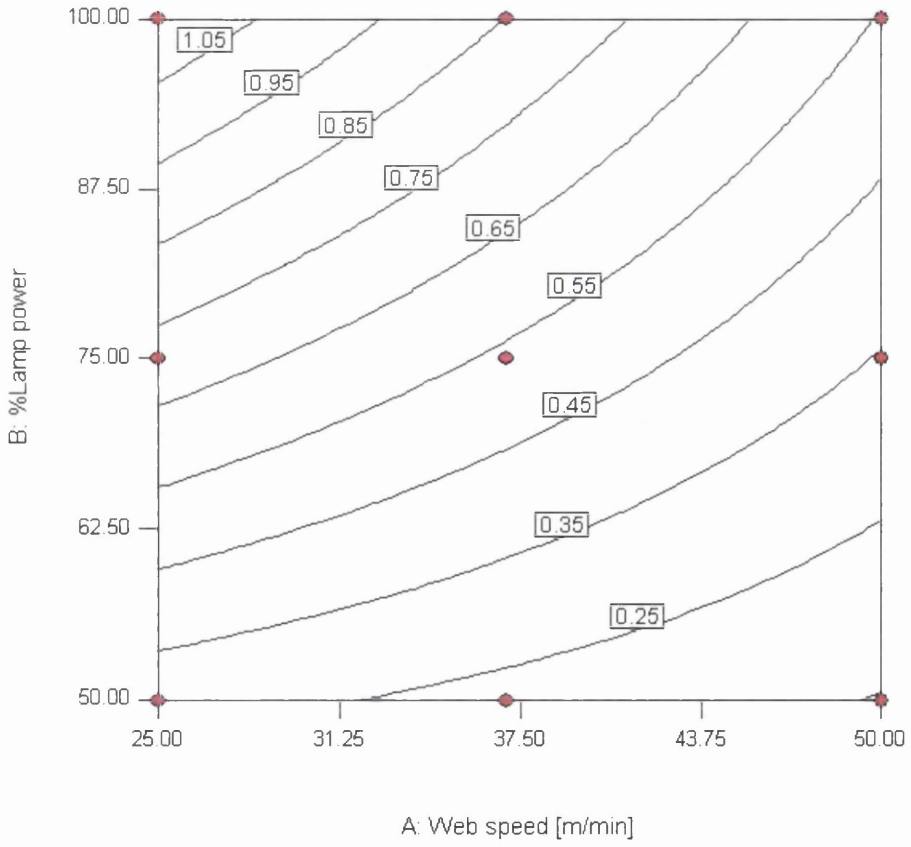


Figure 54 Contour plot for dose prediction generated from the model derived from the central composite DoE.

6.3.4.1.2 Web Temperature

The ANOVA for the dose web temperature shows that a good model can be derived

Source	Sum of Squares	DF	Mean Square	F Value	Prob > F
Model	4056	1	4056	87.5863771	< 0.0001
B	4056	1	4056	87.5863771	< 0.0001
Residual	324.16	7	46.30857		
Cor Total	4380.16	8			
Std. Dev.	6.80504		R-Squared	0.925993571	
Mean	91.46667		Adj R-Squ	0.915421224	
C.V.	7.439913		Pred R-Sq	0.882456794	
PRESS	514.858		Adeq Preci	16.20984674	

Table 18 ANOVA for web temperature

The F Value associated with the model is >87. Prob > F value is 0.0001; therefore the model effect is large. This value implies that the chances of getting an F value this large if the result was due to noise is <0.01%. Therefore the result is almost certainly due to altering the % lamp power.

Surprisingly, line speed over the range investigated, was not significant.

The adjusted R squared value is a measure of the amount of variation about the mean explained by the model using the R-squared multiple correlation coefficient adjusted for the spread of data. The closer to 1 the value, the higher the amount of variation explained by the model. In this case the index is >0.91, therefore the model explains the majority of variation observed during the experiment. The model is described by the following linear equation:-

Equation 6

Web temp °C = 13.7+1.04 * %Lamp power

A graphical representation of this equation is shown Figure 55. The contours are web temperature and the red circles are the conditions run.

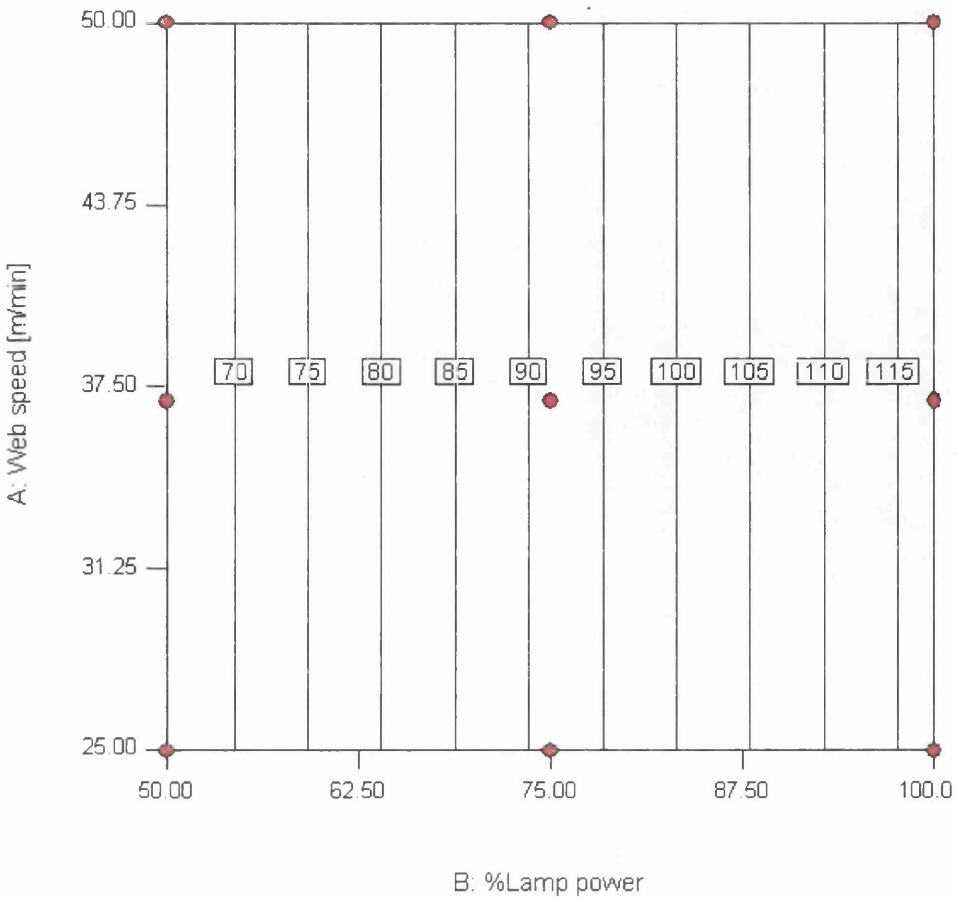


Figure 55 Contour plot for web temperature prediction generated from the model derived from the central composite DoE.

6.3.4.2 The Implications of the findings

The implications of these findings are that for the system to achieve 0.8J/cm², each row of lamps must be set to 82%. This is because the dose from each row of lamps was previously validated to be additive, therefore each lamp must output 0.4J/cm². The contour plot shown below demonstrates the process conditions required for a lamp to output 0.4J/cm² at 50m/min. Running excessively high at say 100% does give a higher web temperature than at 82%. It was not investigated to see how the number of rows of lamps affected web temperature. However, running both rows of lamps at 82% does not give web-handling problems.

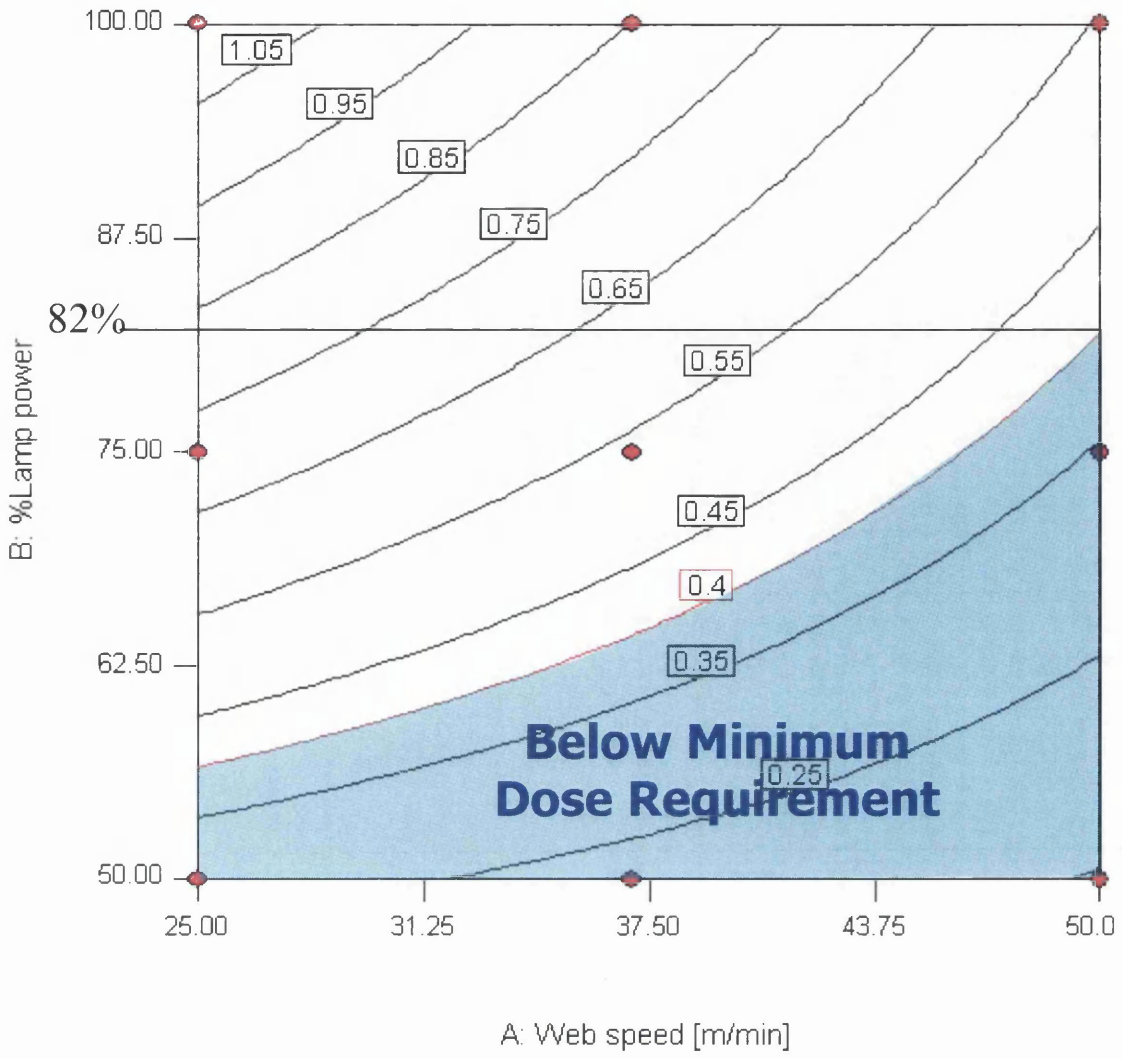


Figure 56 Contour plot of the model showing the process conditions required to achieve the target dose whilst not producing excess UV.

6.4 Closure

Models that describe the operating performance of the feeding and UV curing systems have been developed. These were used to provide inputs to the neural network model of the full-scale system described in the next chapter.

Chapter 7 Modelling the Production System

The pilot-scale and full-scale equipment was used to generate a neural network model to gain process and product under standing.

This chapter starts with a description of the data used for the training the second generation neural network model and how DoE's were used to efficiently cover the large amount of variables and levels from spanning separate processes, followed by the application and results of the neural network modelling methodology laid out in Chapter 4. It concludes by detailing how the scale-up of the processes benefited from using this approach.

7.1 Description of Second Generation Neural Network Model

The datasets for the second generation neural network model are summarised in Figure 17.

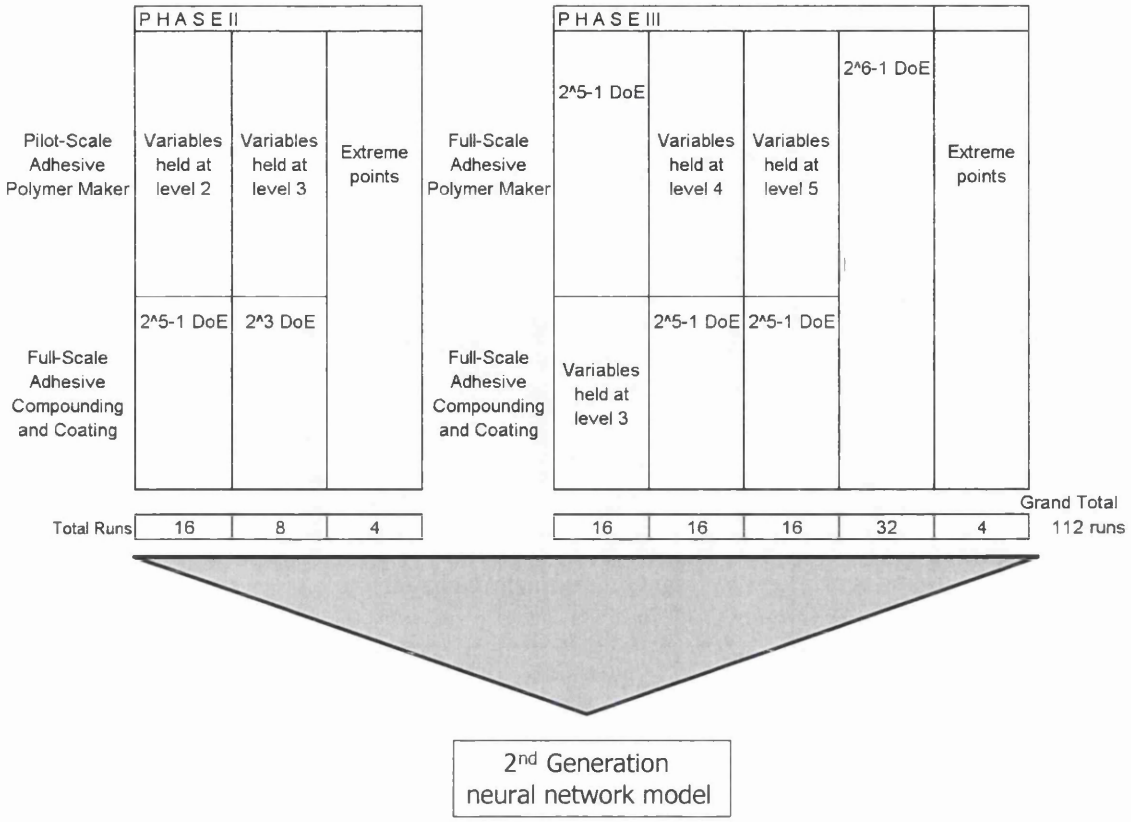


Figure 57 Summary description of data sets that fed the second generation neural network

A further six DoE 's were run that built on the three run to generate the first generation neural network model, described in Chapter 5. Two were run on the full-scale coating and compounding process; a 2^{5-1} fractional factorial design and a 2^3 full factorial design using pilot scale adhesive. Four were run on the full-scale processes, three 2^{5-1} fractional factorial designs and a 2^{6-1}. The other process variables that were not factors for the individual DoE's were also altered both within and between DoE's. Both in phase 2 and phase 3, four "extreme" points were run. These points represented extreme settings of two different parameters of formulation in each phase. These conditions were run to ensure that the neural network was aware of the effects of these formulation parameters when at extreme values. This was to encourage the neural network to interpolate rather than extrapolate whilst predicting tape properties at extreme levels of raw materials. The resulting adhesive polymer was made into tape and tested. The test results and process condition data were combined into a

master database. Some of the material was re-tested after several weeks had elapsed to educate the neural network about any tape ageing phenomena. DoE arrays were used as a means of efficiently covering the large number of variables that could have contributed to tape properties, to generate training and test data for the neural network. Therefore they were not analysed in their own right. This approach, similar to the first generation neural network section, makes use of the fact that DoE arrays are structured, i.e. have high and low levels. An advantage of this approach is that the DoE structures ensure that every point is unique, and therefore of value for modelling.

7.1.1 Application of Neural Network Modelling Methodology

The model was generated from a data matrix using the methodology described in the Chapter 4, and discussed below step by step.

7.1.1.1 Step 1. Generate the modelling data from the total data collected.

All the data generated by the processes was scrutinised and filtered for the following:

a. Highly correlated input variables. Filtering for highly correlated input variables results in discarding all but one representative input. The levels of the discarded variables must be remembered during model interpretation as they are implied, not specified by the neural network model.
b. The categorical data was converted to numerical data.

It was decided to follow the same response-combination strategy for the creation of this second generation neural network model as with the first generation neural network model, i.e. the responses of adhesion and rolling wheel tack were combined in to one response, using a variable column in the data set to instruct the neural network which response value came from which test. This variable column contained a code of 1's or 2's depending on whether the response value came from the adhesion test or the rolling-wheel test. A variable column was also added for which side of the tape the result came from; liner-side or face-side. The effect of combining responses in this manner is that four responses, face-side adhesion, face-side tack, liner-side adhesion and liner-side tack were converted in to one response. This means that the

neural network trained to predict only one response, the combined response. This combined response has analysable subsets within it that can be decoded in order to discern the separate responses. After considering the alternatives to this response-combining strategy, i.e. one neural network predicting four responses or four neural network models each predicting one response, it was deemed more effective to go with the 'one response' option. This meant that the resulting model could be considered as a single virtual process, simultaneously producing an adhesion and a tack property on both the face and liner-side of the virtual product, per settings configuration. If four models were used, four independently configurable virtual processes would each produce an independent adhesion and tack property. Working with four models instead of one multiplies the risk of making mistakes, especially when manipulating data during virtual optimisation experiments. A mistake at this stage could lead to erroneous conclusions and a discredited model. The combined response option was evaluated to ensure it produced a reasonable model.

7.1.1.2 Step 2. Split the data into two data sets, training data and test data.

The data was sub-divided into two sets; training data and test data. Runs, represented by rows in the data in the matrix, were designated in to these two sets by assigning a random number to each row and ordering the matrix by that random number. The first 85% of rows were designated as the training data set and the remainder as test data set. The training data set was used to train the neural network and the test set was used to validate the trained neural network model. Some slight adjustments to the random designations were made to ensure that the range of the responses was similar in both sets. This was done so that testing of the trained neural network model could be done over the widest possible range of responses. The neural network was not exposed to the test set data whilst developing its model, the intention being that predictions made of the test set would not involve recall, and therefore would test to see how well the neural network could predict unseen data. Although the data-generation DoE's had ensured that there was no data that did not represent a boundary of the system, and

each data point was necessary for teaching the neural network about that particular region of design space, it was assumed that the larger volume of data meant that there was more similarity between these unique points than occurred when developing the first generation neural network model. Therefore it was assumed to be less of a risk to deny the neural network access to the information represented by data in the segregated test set during training. There is a risk with this modelling strategy that the model may over-generalise or under-fit the data. The fit of the data was considered in step 4.

7.1.1.3 Step 3. Specify the number of neurons in the hidden layer and define the number of cycles the neural network trains over.

The maximum number of training cycles was set in the software to a number, 150,000, that would ensure a converged solution. The first neural architecture was configured and the network trained. Every 3,000 training cycles the training data set was predicted by the latest network configuration and the RMS error calculated and was automatically logged. This was repeated until the number of training cycles matched the maximum set in the software. This was repeated for each attempt at finding the optimum number of hidden layer neurons. The results were graphed as a plot of RMS error verses the number or training cycles for each network architecture and is shown in Figure 18. There is the typical early rapid decrease in RMS error followed by a tailing off in the rate of reduction.

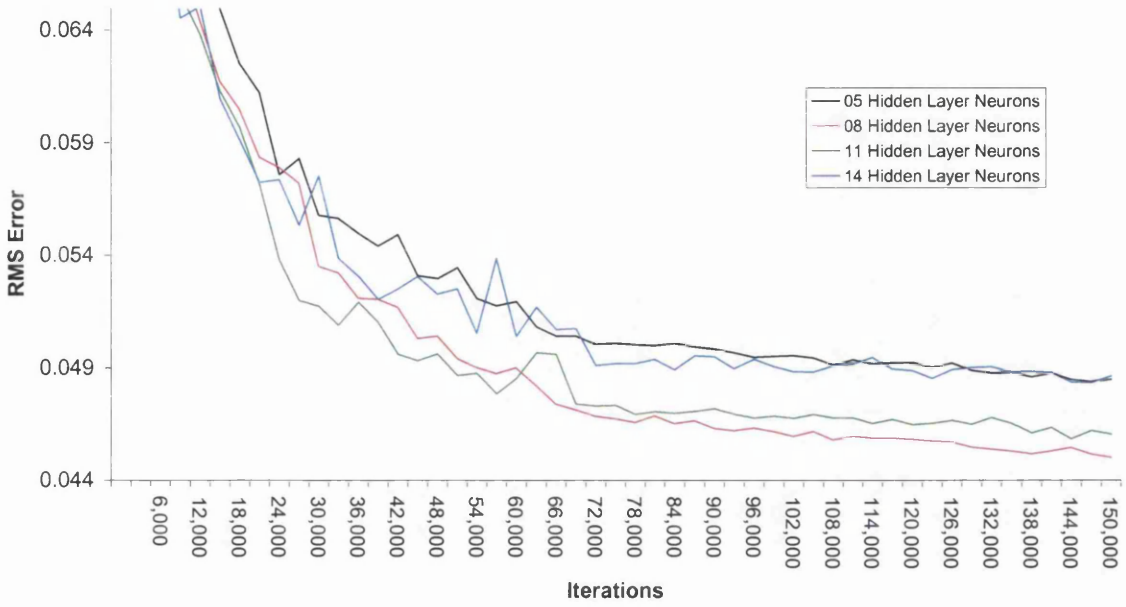


Figure 58 plot of RMS error vs Learning cycles for different numbers of hidden layer neurons

The graph above indicates that the optimum number of hidden layer neurons is 8 (because the lowest RMS error was achieved with this network configuration) and that about 72,000 learning cycles is enough to differentiate the RMS errors (and therefore the accuracies) of the various neural network models resulting from the various architectures. The RMS error seems to be levelling off at 150,000 cycles. To get a modest decrease of RMS error from this point onwards, the number of learning cycles must increase dramatically. The extra benefit of a lower RMS error than that at 150,000 cycles was considered to be negligible. The neural network with 8 hidden layer neurons, trained over 150,000 learning cycles, was preserved and became the second generation neural network model. The resulting model was checked for over-fitting.

7.1.1.4 Step 4. Check the chosen model for over-fitting.

Predictions of the neural network training data set were compared with the actual values using an XY plot, shown in Figure 59. This graph is a plot of the model's

predictions for the combined response vs. the actual measured values, with a line of ideal fit through the data. This line represents where the data should fall if the predictions were perfect. The graph shows two things. The first is general agreement between the actual and predicted values. If there is, there will be a correlation between the data, i.e. the bulk of the data will lie on the line of ideal fit within a certain error band (the adjusted R^2 for this data is 0.809). The second is to raise suspicion if the model has over-learned the training data, i.e. the neural network has modelled noise. This would be highlighted if the majority of data rarely deviated from the line of ideal fit. A quantitative assessment would then be required to confirm if the prediction error is less than the historical test/ retest error for that measurement system. If it was then overfitting can be assumed to have occurred.

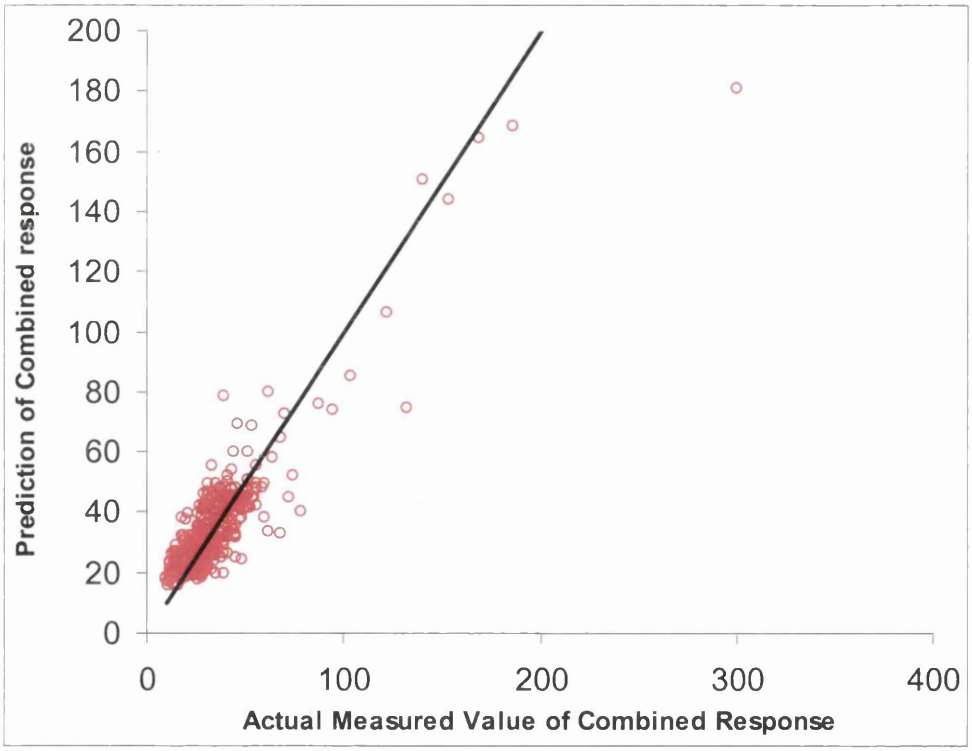


Figure 59 Correlation plots of predicted and actual measurements for the combined responses in the training data set.

The model's predictions of the combined response seem to broadly follow the underlying trend of the actual values, albeit with some error. This is confirmed up by the ANOVA for the regression shown in Table 19 showing that there is significant

agreement. Therefore the neural network has learned some true relationships between the variables and responses. At the same time there is some error in the prediction. The neural network has therefore not over-learned the data and has not over-modelled noise. It has been able to generalise.

Source	Sum of Squares	DF	Mean Square	F Value	Prob>F
Regression	140158	1	140158	3605.34	0.000
Residual Error	33122	852	39		
Total	173280	853			

Table 19 ANOVA for the linear regression of predicted vs actual combines response

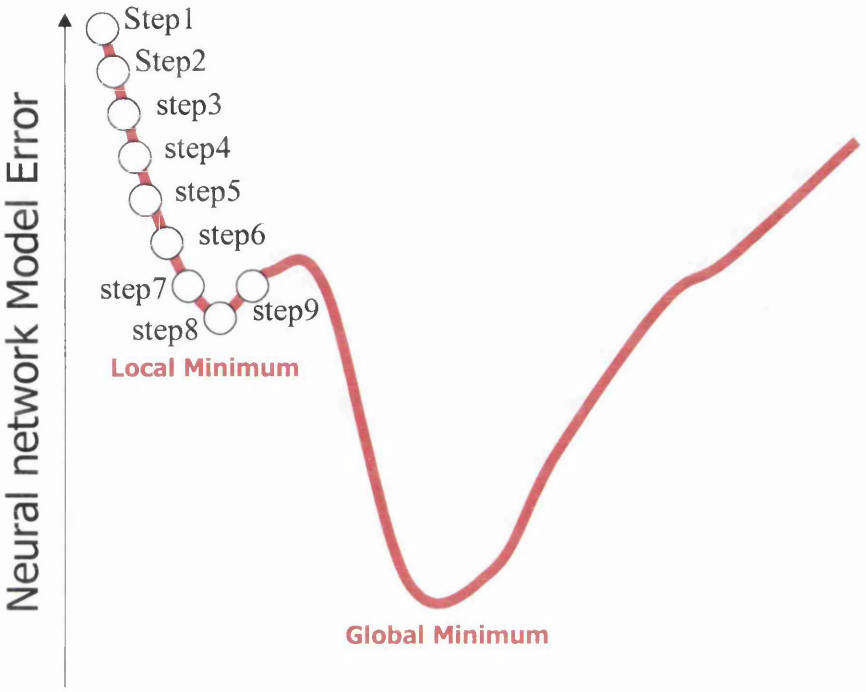
The level of generalisation at this stage was of some concern. There was variation not explained by the model, about 19% according to the adjusted R^2 index calculated to be 0.809. This concern caused the inclusion of an extra step of model validation that was not required with the first generation neural network, described in Chapter 5.

7.1.1.5 Step 4a. Neural network model validation.

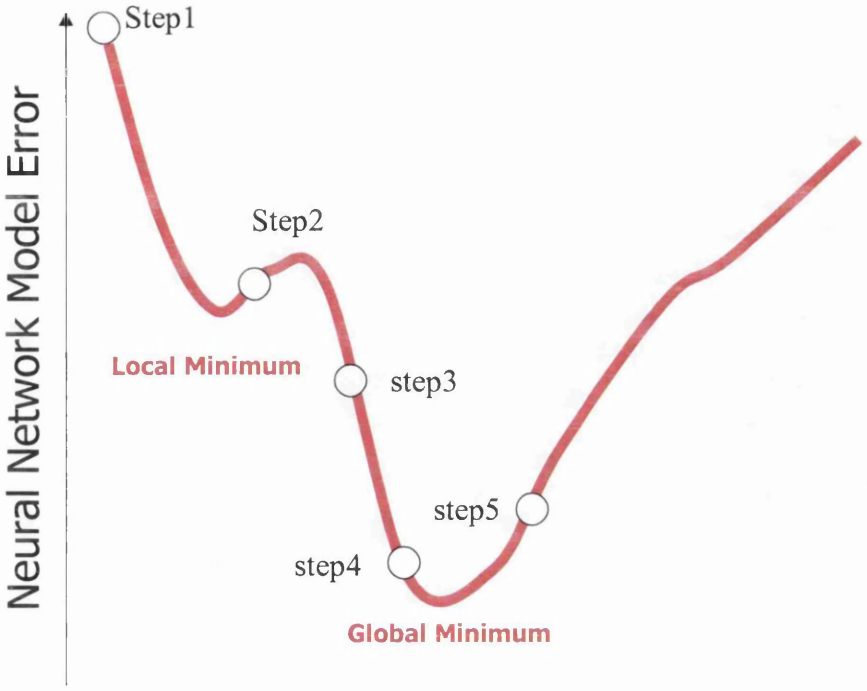
The prediction error level of the neural network model may be this high due to some correctible deficiency. If this were so, it would mean that the model had sub-optimal predictive properties. Poor predictive performance would reduce confidence in the relevance of any recommendations derived from a model. Therefore an investigation in to the cause of the apparently high level of prediction error was performed to see if it could be corrected. The main reasons why a neural network model does not predict as well as it could are listed below.

1. Sub-optimal learning parameters used. The neural network enhances its model during training by minimising the error between the actual data and its predictions during each learning cycle, using a stepwise approach. A step-size that was too small may have resulted in the neural network stopping learning

prematurely at some local minimum error rather than the global minimum error, as in the representation below.



In this representation, during training, the control strategy would have forced the neural network to home in on to the conditions around step 8 after realising that going to step 9 was detrimental to the level of model error. The remedy to avoid stopping learning at a local minimum is to increase the step size such that the local minimum is by-passed, as in the representation below.



Here the neural network control strategy would have forced the neural network to home in on the conditions around step 4 after realising that going to step 5 was detrimental to the error of the model.

2. There is a variable acting on the system that the neural network is unaware of. This would cause tangible unpredictability depending on the strength of its effect. The remedy for this would be to search for this "mystery" variable by revisiting the choice of variables included, in the neural network model. For new variables to be included and the data already exists in the master database, then this is a simple case of retraining the neural network with the new variables included. If on the other hand no data exists for new variables, this could invalidate all the data collected.
3. Part of the data is affected by a mystery variable or is partly corrupted. The remedy is to search for this mystery variable or the corruption, and remove it from the existing model.

Investigation in to reason 3 was carried out after considering the remedies for the other two reasons. Reason 1 was considered to be the least likely of the three because

the learning strategy had been somewhat optimised during other modelling projects[17]. Remedial action for reason 2 would only be considered after rejecting reason 3 as the source of the model error. Reason 3 was therefore investigated first.

Residuals are repeatedly used in this investigation exercise, to test some of the basic assumptions about a model. These are that if the model fits the data correctly, then any error (residuals) would be the result of normally distributed random effects such as test error, experimental noise etc. If the residuals are shown to be random and normally distributed, then it implies that no mystery process variables are acting. In which case if the model is required to be more precise, effort must be directed towards reducing test error, experimental noise, etc. If the residuals are shown not to be random or normally distributed, then the influence of a mystery variable can be assumed to be at work quite legitimately.

The search for the influence of a variable acting on part of the data began by testing the normality of the residuals. Normal probability plots highlight if the data is normal (Figure 60). Normally distributed data would all be shown in a straight line. As the data does not all fall in a straight line at the extremes, significant departure from normality can be assumed. This implies that a mystery variable has influenced the data.

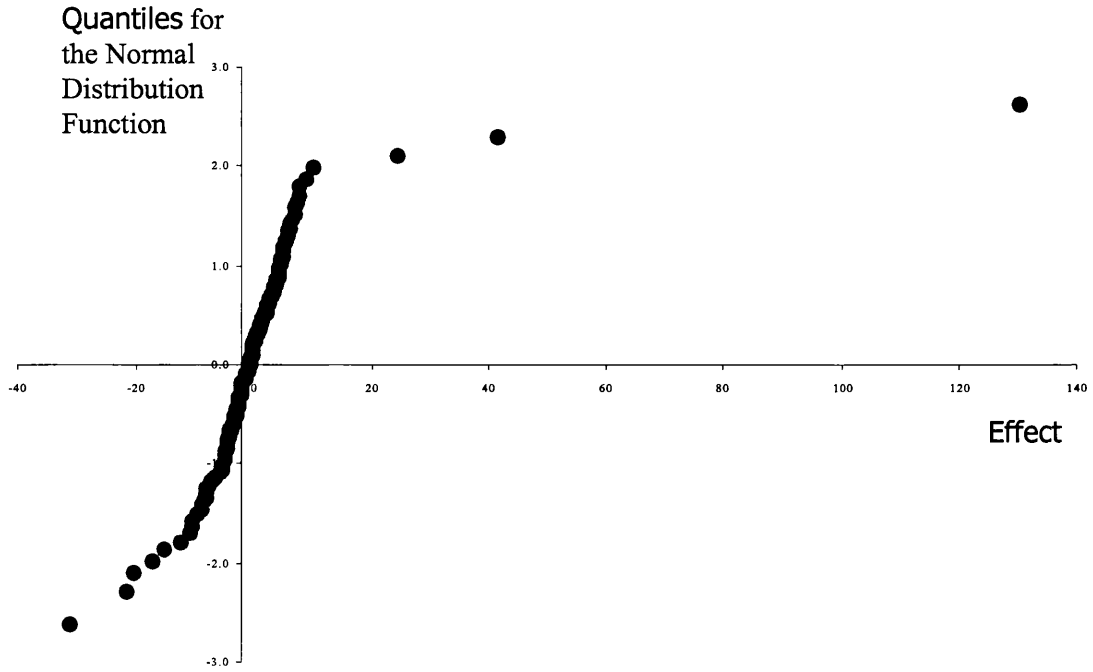


Figure 60 Normal probability plot of the residuals of the neural network model predicting the test data set.

Having shown that a significant mystery variable probably exists using normal probability plots, the search for this mystery variable began by testing the assumptions made in step 1, i.e. that combining responses in to one response had not caused data corruption or manifested the mystery variable. The residuals (from test data set) for each of the four subsets of the combined response were plotted on a dotplot (Figure 61). The four subsets of the combined response were the four possible combinations of the following two choices:-

a. side of tape tested (face or liner side) and

b. the actual test carried out (adhesion or rolling wheel tack).

A dotplot displays a dot for each value along a number line. If there are multiple occurrences of the same response value, or if response values are close together, then dots are stacked vertically. This is similar to a histogram. This type of plot allows one to assess the nature of the spread of data, whether there are any outliers and where the data is concentrated.

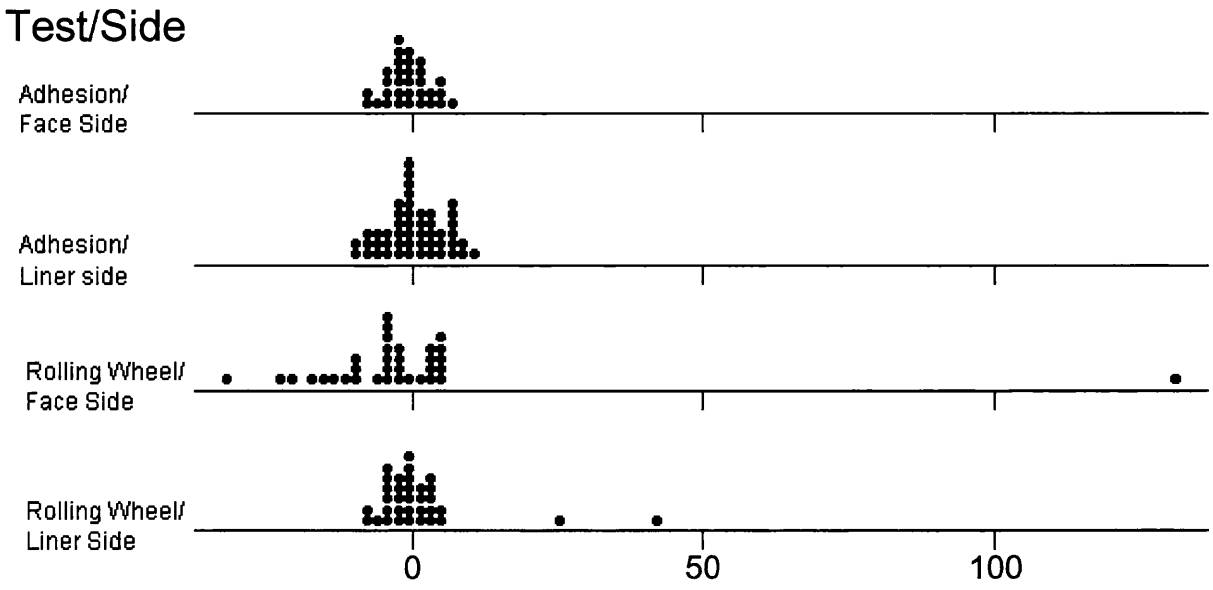


Figure 61 Dotplot of the residuals of the trained neural network model predicting the test dataset for each subset of the combined response.

The adhesion residuals and the liner-side rolling-wheel residuals are grouped around zero, have similar spreads and approximate normal distributions. There are some outliers in both rolling-wheel populations. This visualisation of the data hints that the mystery variable is acting on the face-side rolling-wheel results only, all other spreads of residuals look like they come from the same population, with the same variance (standard deviation^2) and with the same mean. To test this hypothesis, the face-side rolling wheel data was removed from the test set, together with the two outliers from the liner-side rolling-wheel data and this modified test data set was rechecked for normality using the normal probability plot approach. The plot is shown in Figure 62.

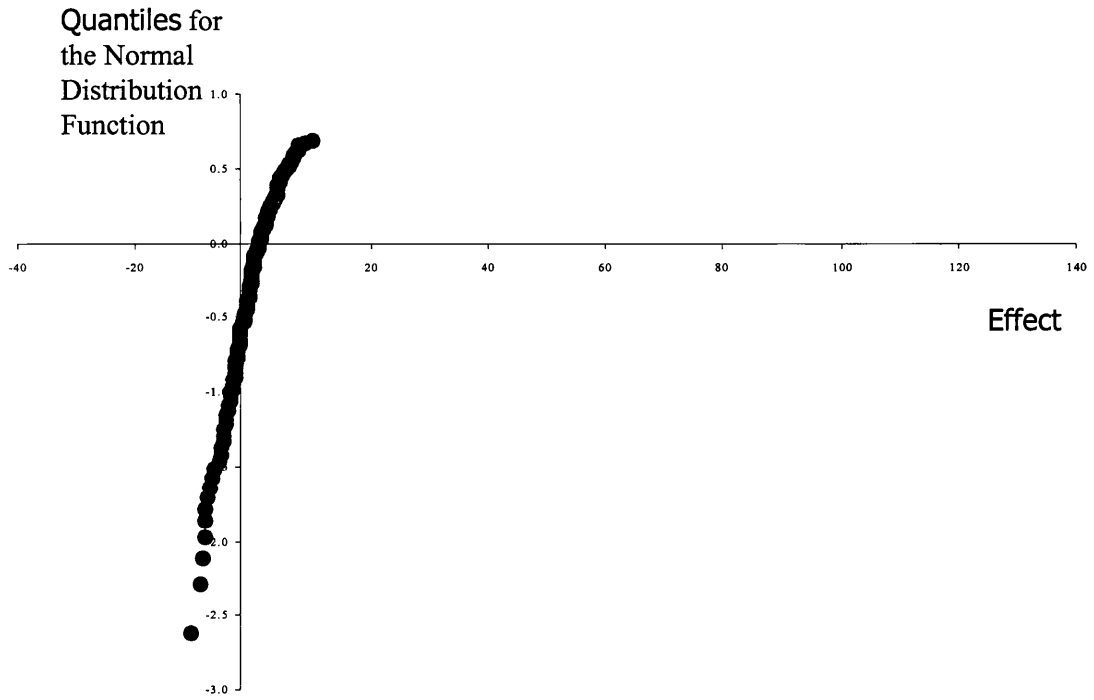


Figure 62 Normal probability plot of the residuals of the neural network model predicting the test data set with the face-side rolling wheel data and outlier data removed.

The data is more normal, i.e. the line is straighter. Normal probability plots are only a graphical assessment of normality. A numeric check was carried out using the Anderson-Darling Normality Test. This test returns an index, A^2, that has an associated probability or P-value of being at that level if the data comes from a normal distribution. In this case A^2 is 0.214 giving a corresponding P-value of 0.848. This confirms that the data is probably normally distributed and therefore the mystery variable was only influencing the face-side rolling wheel tack results.

As a further confirmation, this modified test data set should have residuals centred on zero, i.e. if random noise is the only source of error in the model, then the effect of this noise is just as likely to add to the combined response as subtract from it. This was proved by carrying out a T-test on the residuals of the prediction of the combined response (not including face-side rolling-wheel results). This is summarised in the

histogram in Figure 63. The fact that the hypothesised mean, zero, falls within the 95% confidence interval of the mean confirms that the residuals are centred around zero.

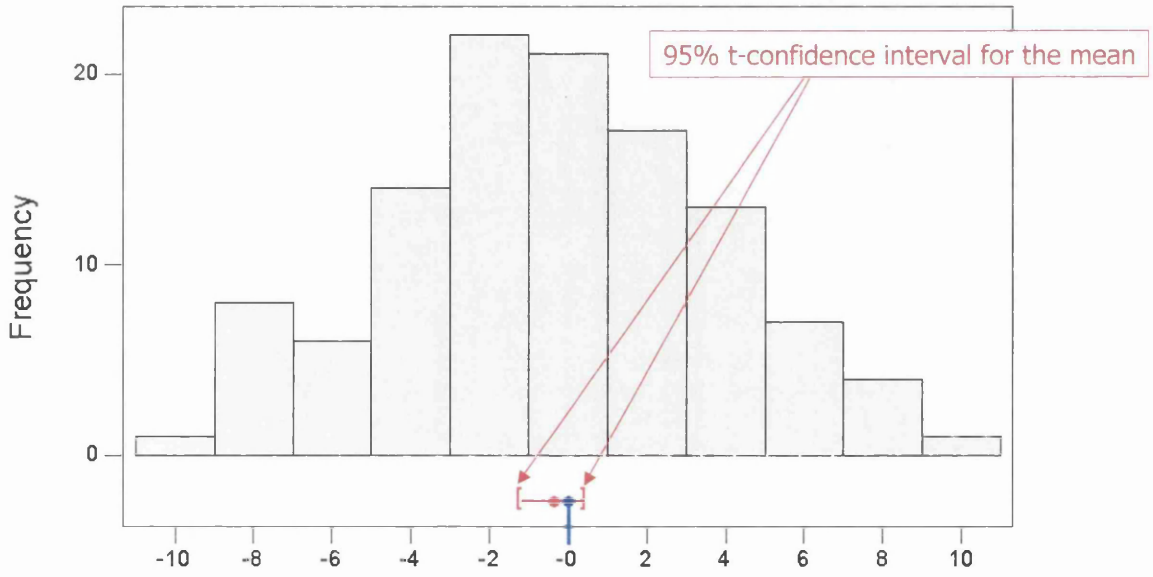


Figure 63 Histogram of the residuals of the test data set with face-side rolling-wheel and outlier data removed, with the 95% t-confidence interval of the mean.

If the residuals are truly random, then the variances of the residuals from each subset would be the same and independent of test type and side tested. A statistical method for checking the difference between the spreads of different subsets of data is the Bonferroni test. This test allows one to calculate the confidence interval of the standard deviation of each subset. Overlaying the confidence intervals on a number line allows comparison (Figure 64). Where there is overlap, the variances have a certain probability of being drawn from the same population. In this case a 95% probability.

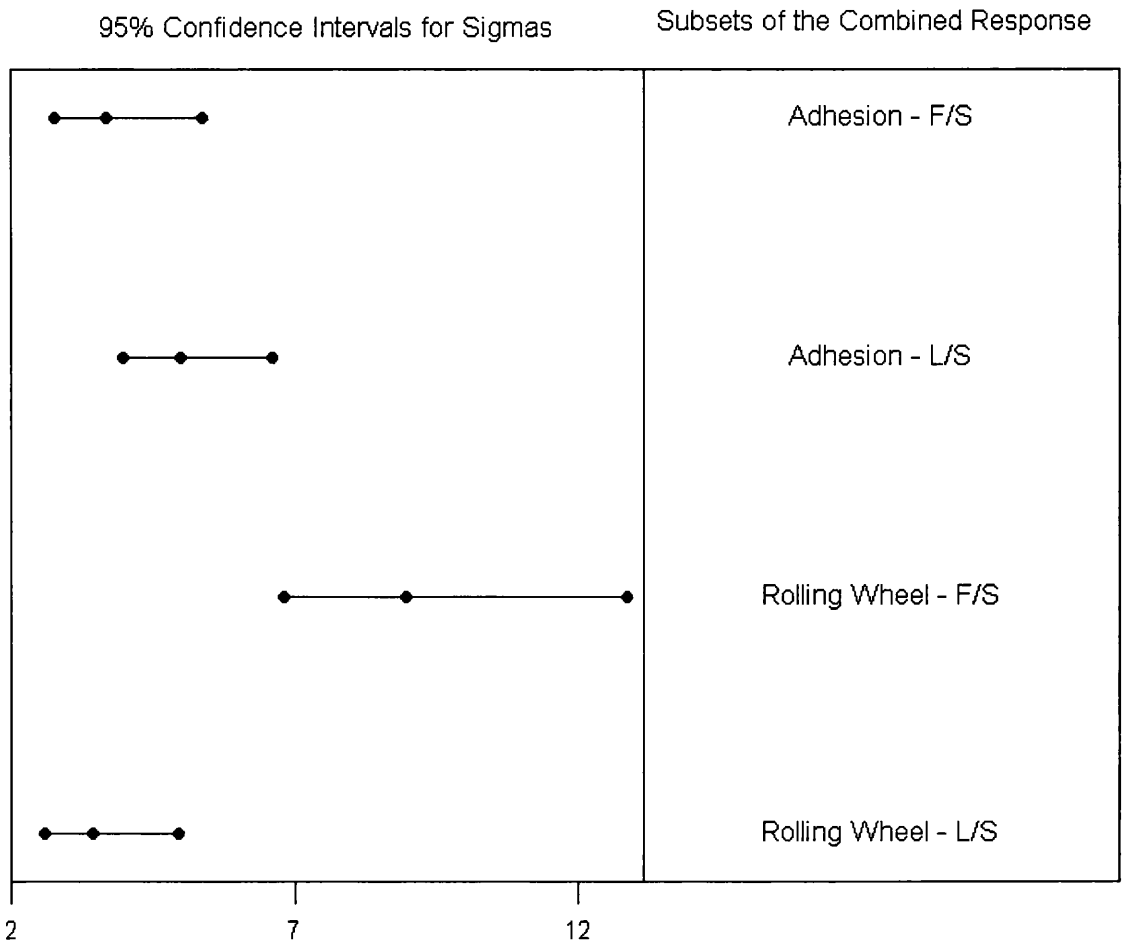


Figure 64 95% Confidence intervals of the standard deviations for the subsets of the combined response.

It is likely that the residuals of the subsets are from the same populations except those from the face-side rolling-wheel data set (Figure 64). This is further justification for rejecting the data from this subset because of a mystery variable influencing the results.

Further evidence that the remaining subsets' residuals are the result of normal randomness, and therefore not as a result of a mystery variable, is that the magnitude of the residuals is independent of the level of the prediction. This was checked by plotting the residuals against the predicted value and checking the consistency of the spread (Figure 65). If the residuals were dependent on the level of prediction, a

tapering pattern of spread would exist across some or all of the data. If there were independence, no tapering would exist.

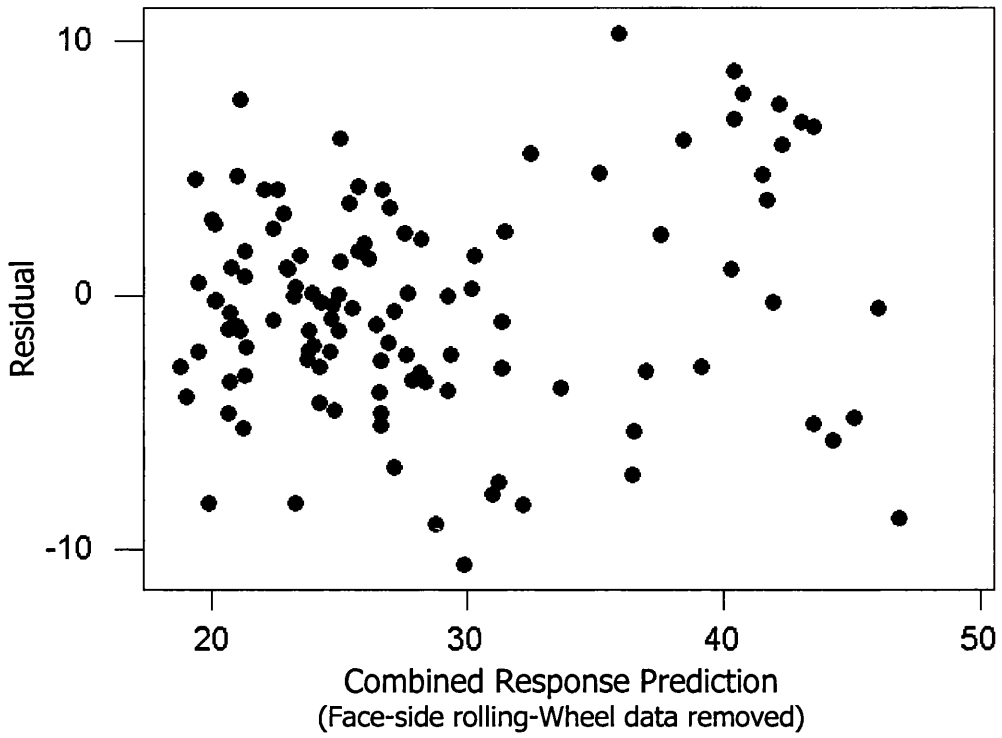


Figure 65 Plot of the residuals verses combined response level.

The residuals are evenly distributed across the whole range of values of combined response (Figure 65). This further justifies the assumption that the model has correctly predicted the underlying test set data, and the remaining error is down to test error, process noise etc.

7.1.1.6 Step 5. Interrogate the neural network model and interpret the results.

The purpose for generating this second generation neural network model of the pilot-scale and full-scale processes was to gain process and product understanding. This would enable rapid optimisation of the full-scale equipment, to facilitate rapid transfer of products between the pilot-scale and full-scale processes and provide a predictive quality tool, and to aid the search for product opportunities such as "super-properties".

The trained neural network model is a "virtual" process that can be run without having to make real product. As such, one way to assess if there are any interesting phenomena, or if optimised properties exist, is to run a DoE on the virtual process and make use of response visualisation techniques such as contour plots and response surfaces. Because the process is virtual, and no real product has to be made, large DoE's can be run. The size of the DoE's or number of questions that can be run on virtual processes is only limited by the time it takes to prepare the data. This is measured in minutes rather than days for real DoE's.

7.1.1.6.1 The relationship between adhesion & rolling-wheel tack

Before analysing the central composite DoE, the virtual responses of liner-side adhesion and rolling-wheel tack were plotted against each other (Figure 66).

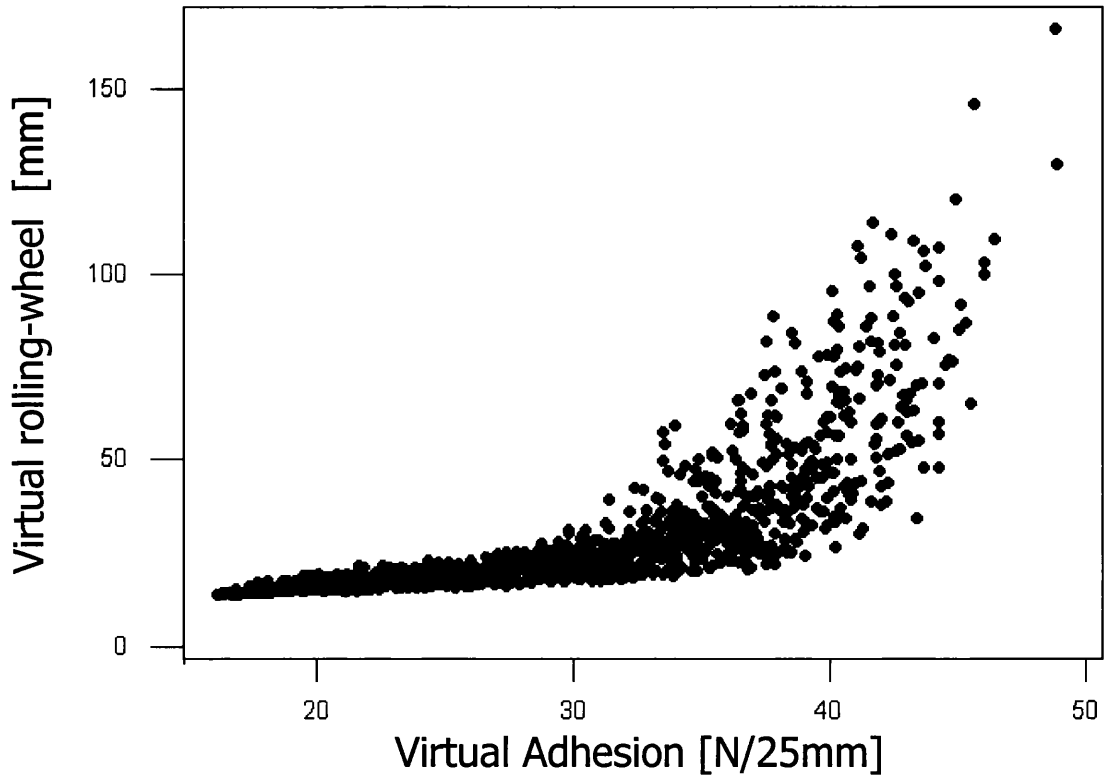


Figure 66 XY Plot of the virtual responses (liner-side only) for the central composite DoE data

There appears to be dependence between the two responses, i.e. the data has a discernable pattern and is not randomly distributed. It also appears that adhesion is inversely proportional to rolling-wheel tack. This was tested by performing linear regression on the relationship between virtual adhesion and 1/(virtual rolling-wheel), (Figure 67).

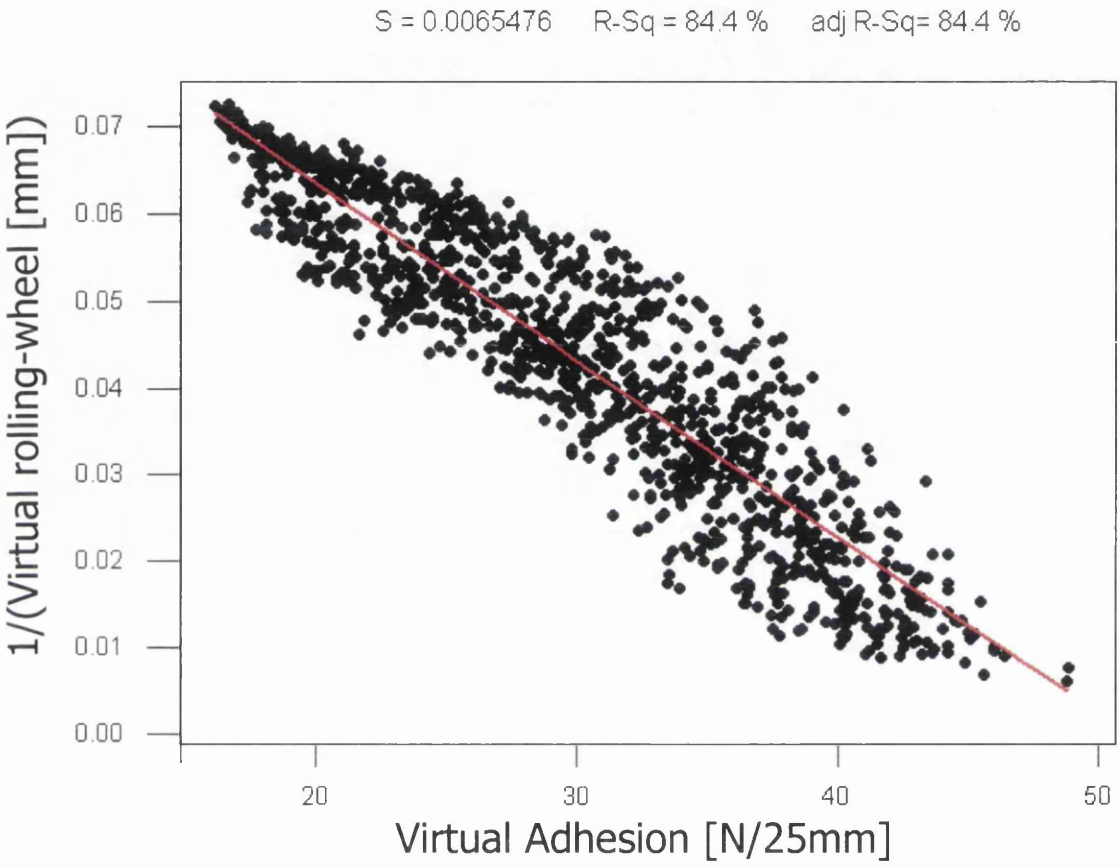


Figure 67 XY plot showing the degree of correlation between virtual adhesion and reciprocal virtual rolling-wheel data generated for the central composite DoE

The linear regression yielded the following equation for the relationship:-

$$\textit{virtual rolling-wheel tack} = 1/(-0.002\ \textit{virtual adhesion} + 0.1041)$$

The regression analysis had the following significant ANOVA.

Source	Sum of Squares	DF	Mean Square	F-Value	Prob >F
Regression	0.29274	1	0.29274	6828.39	0
Error	0.054103	1262	0.000043		
Total	0.346843	1263			

Figure 68 ANOVA of linear regression of inverse tack and adhesion

The regression is therefore valid and the majority of the variation of virtual rolling-wheel is explained by the level of virtual adhesion; adjusted R^2 is 0.844. A test of the magnitude of the residuals, in the form of a plot of the residuals against the predicted value, was made to reveal the level of randomness, hence indicating response dependency (Figure 69). If the residuals were dependent on the level of prediction, a tapering pattern of spread would exist across some or all of the data. If there is independence, no tapering pattern would exist.

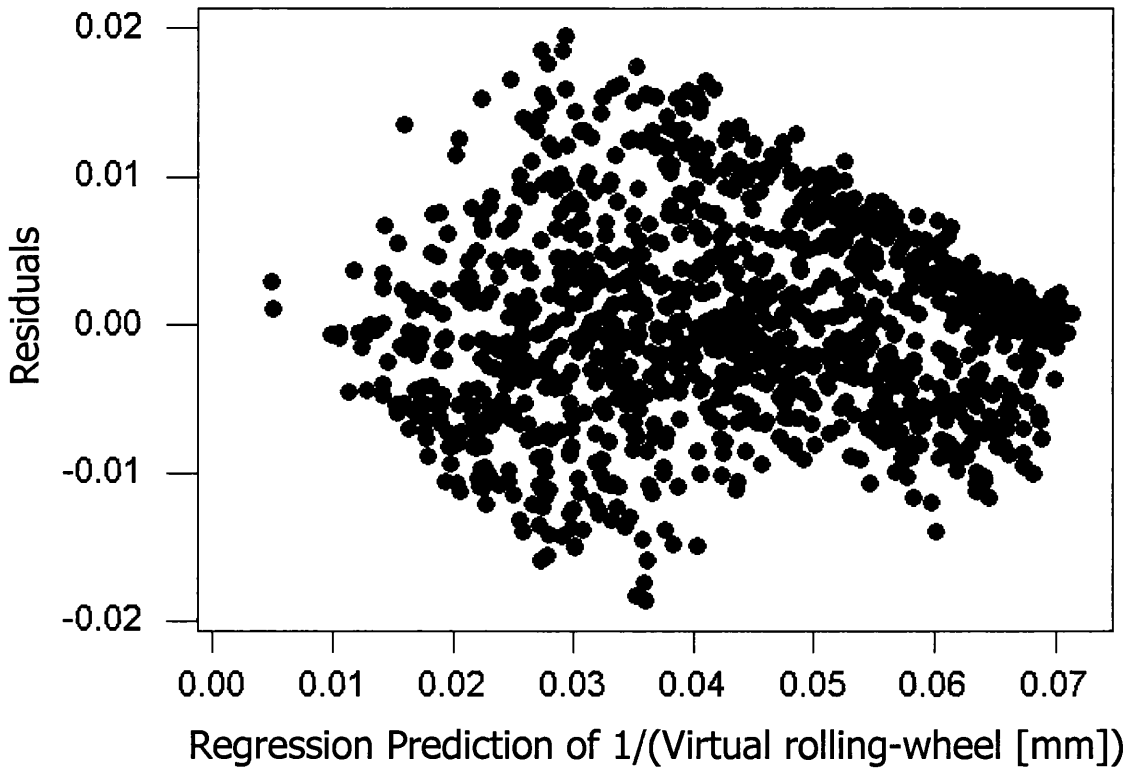


Figure 69 XY plot of the residuals from the linear equation verses the prediction of virtual reciprocal rolling-wheel.

There is shape to the distribution of residuals and therefore there is departure from the dependency of the two virtual responses. Therefore virtual adhesion is not a constant predictor of virtual rolling-wheel tack. However the linear equation should be sufficient if only approximate values are needed.

7.1.1.6.2 Overview of the DoE used to interrogate the neural network model and some general findings.

The first generation neural network model of the pilot processes was interrogated using a low resolution Taguchi L54 DoE, a design in keeping with the level of understanding required at that stage (Chapter 5). In essence a DoE was run on a virtual process making virtual product with properties predicted by the neural network model. The interrogation of the second generation neural network model was pursued using a similar strategy. However, because more detailed understanding was required to answer optimisation questions, a central composite DoE was chosen to interrogate the second generation neural network with instead of a Taguchi DoE. This type of design is capable of modelling with quadratic equations and requires five levels of each numeric factor to be run. Importantly the knowledge locked up in the neural network can drawn out and visualised using contour plots and response surfaces, both a part of central composite DoE analytical methods.

In order to get the maximum understanding with the DoE, all the neural network model variables were to be included as DoE factors. However, the available analysis software is limited to the number of factors that could be simultaneously studied. This is presumably in line with the assumption that it is too costly or inappropriate to examine more than ten numeric variables at once on most real processes. Constrained by this limitation, the maximum number of numeric DoE factors, 10, was assigned to selected continuous neural network model variables. These were chosen on the basis that they either had a strong likelihood of influencing tape properties or process economics did not require them to be fixed at lowest-cost settings. The DoE factors and design space covered by this central composite DoE are summarised in the table below.

Coded Factor Name	Factor Name	- level	+ level
A	Adhesive membrane thickness	2.24	2.54
B	Water Bath average Temp	58	85
C	UV Power 1st zone	1.0001	4.53
D	Total Radiation	1808.8	3946.7
E	Reaction Time	493	1509
F	Age of adhesive at time of coating	37	261
G	Resin A : Adhesive Polymer	15.008	61.18686
H	Resin B : Adhesive Polymer	4.9966	25
I	Resin C : Adhesive Polymer	0.1497	0.717447
J	Age of Tape at testing	0	71
K	Pilot-Scale or Full-scale Adhesive Polymer maker (1 or 2)	1	2
L	Pilot-Scale or Full-scale Compounding or Coating (1 or 2)	1	2
M	Who Tested, laboratory 1 or 2	1	2
N	Side of tape tested, Face or Liner (1 or 2)	1	2
O	Test, Adhesion or Rolling-Wheel tack (1 or 2)	1	2

Table 20 Factor codes, names and levels

The neural network predicted the virtual product properties of adhesion and rolling-wheel tack. NB, the neural network model of face-side rolling-wheel tack was shown to be unreliable, therefore DoE analysis did not involve considering this portion of the combined response modelled by the second generation neural network. This DoE required over 2,500 runs. The DoE was analysed, the statistically significant terms chosen and the ANOVA calculated. The ANOVA tables for virtual adhesion and virtual liner-side rolling-wheel tack shown in Table 21 and Table 22.

Source	Sum of Squares	DF	Mean Square	F-Value	Prob > F
Model	202116.587	50.000	4042.332	900.924	< 0.0001
A	2267.503	1.000	2267.503	505.364	< 0.0001
B	54.030	1.000	54.030	12.042	0.0005
C	516.130	1.000	516.130	115.031	< 0.0001
D	66.299	1.000	66.299	14.776	0.0001
E	34.360	1.000	34.360	7.658	0.0057
F	5275.915	1.000	5275.915	1175.856	< 0.0001
G	76907.841	1.000	76907.841	17140.639	< 0.0001
H	18460.439	1.000	18460.439	4114.323	< 0.0001
J	22.653	1.000	22.653	5.049	0.0247
K	41.462	1.000	41.462	9.241	0.0024
L	66537.897	1.000	66537.897	14829.464	< 0.0001
M	3915.922	1.000	3915.922	872.751	< 0.0001
N	12843.360	1.000	12843.360	2862.431	< 0.0001
O	133.206	1.000	133.206	29.688	< 0.0001
G2	2305.911	1.000	2305.911	513.924	< 0.0001
H2	163.816	1.000	163.816	36.510	< 0.0001
AG	224.556	1.000	224.556	50.047	< 0.0001
AH	80.716	1.000	80.716	17.989	< 0.0001
AL	139.713	1.000	139.713	31.138	< 0.0001
AO	236.543	1.000	236.543	52.719	< 0.0001
BC	20.884	1.000	20.884	4.655	0.0311
BG	31.080	1.000	31.080	6.927	0.0085
BL	24.044	1.000	24.044	5.359	0.0207
BO	49.938	1.000	49.938	11.130	0.0009
CG	28.716	1.000	28.716	6.400	0.0115
EG	105.539	1.000	105.539	23.522	< 0.0001
EH	29.115	1.000	29.115	6.489	0.0109
EL	106.348	1.000	106.348	23.702	< 0.0001
EM	21.016	1.000	21.016	4.684	0.0305
EN	17.465	1.000	17.465	3.892	0.0486
EO	312.065	1.000	312.065	69.551	< 0.0001
FN	23.061	1.000	23.061	5.140	0.0235
FO	181.122	1.000	181.122	40.367	< 0.0001
GH	1203.179	1.000	1203.179	268.155	< 0.0001
GJ	21.646	1.000	21.646	4.824	0.0282
GL	1451.378	1.000	1451.378	323.472	< 0.0001
GN	179.012	1.000	179.012	39.897	< 0.0001
GO	3530.771	1.000	3530.771	786.912	< 0.0001
HL	822.153	1.000	822.153	183.235	< 0.0001
HN	130.862	1.000	130.862	29.166	< 0.0001
HO	1208.803	1.000	1208.803	269.409	< 0.0001
JL	18.307	1.000	18.307	4.080	0.0435
KM	35.453	1.000	35.453	7.901	0.0050
KN	61.049	1.000	61.049	13.606	0.0002
KO	18.976	1.000	18.976	4.229	0.0398
LN	102.188	1.000	102.188	22.775	< 0.0001
LO	1624.772	1.000	1624.772	362.117	< 0.0001
MN	187.351	1.000	187.351	41.755	< 0.0001
MO	125.470	1.000	125.470	27.964	< 0.0001
NO	286.205	1.000	286.205	63.787	< 0.0001
			R-Squared	0.948	
			Adj R-Squared	0.947	
			Pred R-Squared	0.945	
			Adeq Precision	212.346	

Table 21 ANOVA for virtual adhesion derived form the interrogating central composite DoE run on the second generation neural network model.

.

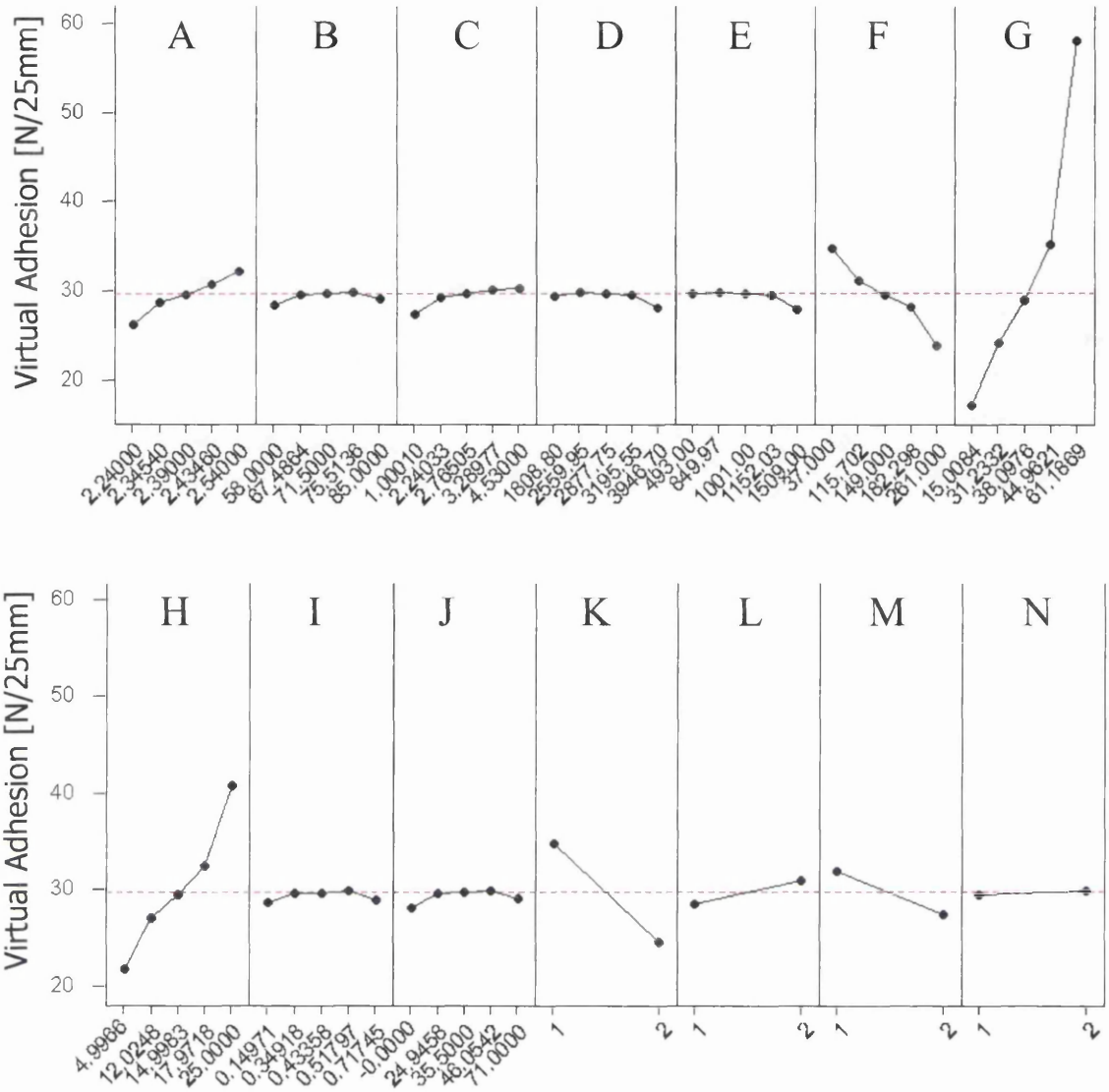
Source	Sum of Squares	DF	Mean Square	F-Value	Prob > F
Model	436414.337	46.000	9487.268	720.953	< 0.0001
A	8260.078	1.000	8260.078	627.697	< 0.0001
B	108.972	1.000	108.972	8.281	0.0041
C	232.217	1.000	232.217	17.647	< 0.0001
E	1234.464	1.000	1234.464	93.809	< 0.0001
F	3038.403	1.000	3038.403	230.893	< 0.0001
G	130746.715	1.000	130746.715	9935.659	< 0.0001
H	29639.733	1.000	29639.733	2252.372	< 0.0001
K	563.711	1.000	563.711	42.837	< 0.0001
L	93284.948	1.000	93284.948	7088.877	< 0.0001
M	38382.878	1.000	38382.878	2916.778	< 0.0001
N	11371.799	1.000	11371.799	864.162	< 0.0001
A2	60.510	1.000	60.510	4.598	0.0322
G2	19390.862	1.000	19390.862	1473.544	< 0.0001
H2	1247.883	1.000	1247.883	94.829	< 0.0001
AG	2765.726	1.000	2765.726	210.172	< 0.0001
AH	679.955	1.000	679.955	51.671	< 0.0001
AL	2470.629	1.000	2470.629	187.747	< 0.0001
AM	554.119	1.000	554.119	42.108	< 0.0001
AN	179.579	1.000	179.579	13.647	0.0002
BC	258.485	1.000	258.485	19.643	< 0.0001
BG	90.080	1.000	90.080	6.845	0.0090
BL	52.848	1.000	52.848	4.016	0.0453
CG	134.291	1.000	134.291	10.205	0.0014
CL	114.144	1.000	114.144	8.674	0.0033
EG	522.259	1.000	522.259	39.687	< 0.0001
EH	176.953	1.000	176.953	13.447	0.0003
EL	400.211	1.000	400.211	30.413	< 0.0001
EM	51.223	1.000	51.223	3.893	0.0487
FG	325.002	1.000	325.002	24.697	< 0.0001
FL	351.077	1.000	351.077	26.679	< 0.0001
FM	267.967	1.000	267.967	20.363	< 0.0001
GH	11456.357	1.000	11456.357	870.587	< 0.0001
GK	308.549	1.000	308.549	23.447	< 0.0001
GL	42095.881	1.000	42095.881	3198.935	< 0.0001
GM	8636.131	1.000	8636.131	656.274	< 0.0001
GN	2991.109	1.000	2991.109	227.299	< 0.0001
HK	68.694	1.000	68.694	5.220	0.0225
HL	11033.608	1.000	11033.608	838.462	< 0.0001
HM	1603.780	1.000	1603.780	121.874	< 0.0001
HN	479.825	1.000	479.825	36.463	< 0.0001
KL	299.481	1.000	299.481	22.758	< 0.0001
KM	68.365	1.000	68.365	5.195	0.0228
KN	61.191	1.000	61.191	4.650	0.0312
LM	7011.729	1.000	7011.729	532.833	< 0.0001
LN	2653.397	1.000	2653.397	201.636	< 0.0001
MN	1322.435	1.000	1322.435	100.494	< 0.0001
Residual	16014.917	1217.000	13.159		
Lack of Fit	16014.917	1145.000	13.987		
Pure Error	0.000	72.000	0.000		
Cor Total	452429.254	1263.000			

R-Squared	0.964602
Adj R-Squared	0.963264
Pred R-Squared	0.960636
Adeq Precision	204.8939

Table 22 ANOVA for virtual liner-side rolling-wheel tack derived from the interrogating central composite DoE run on the second generation neural network

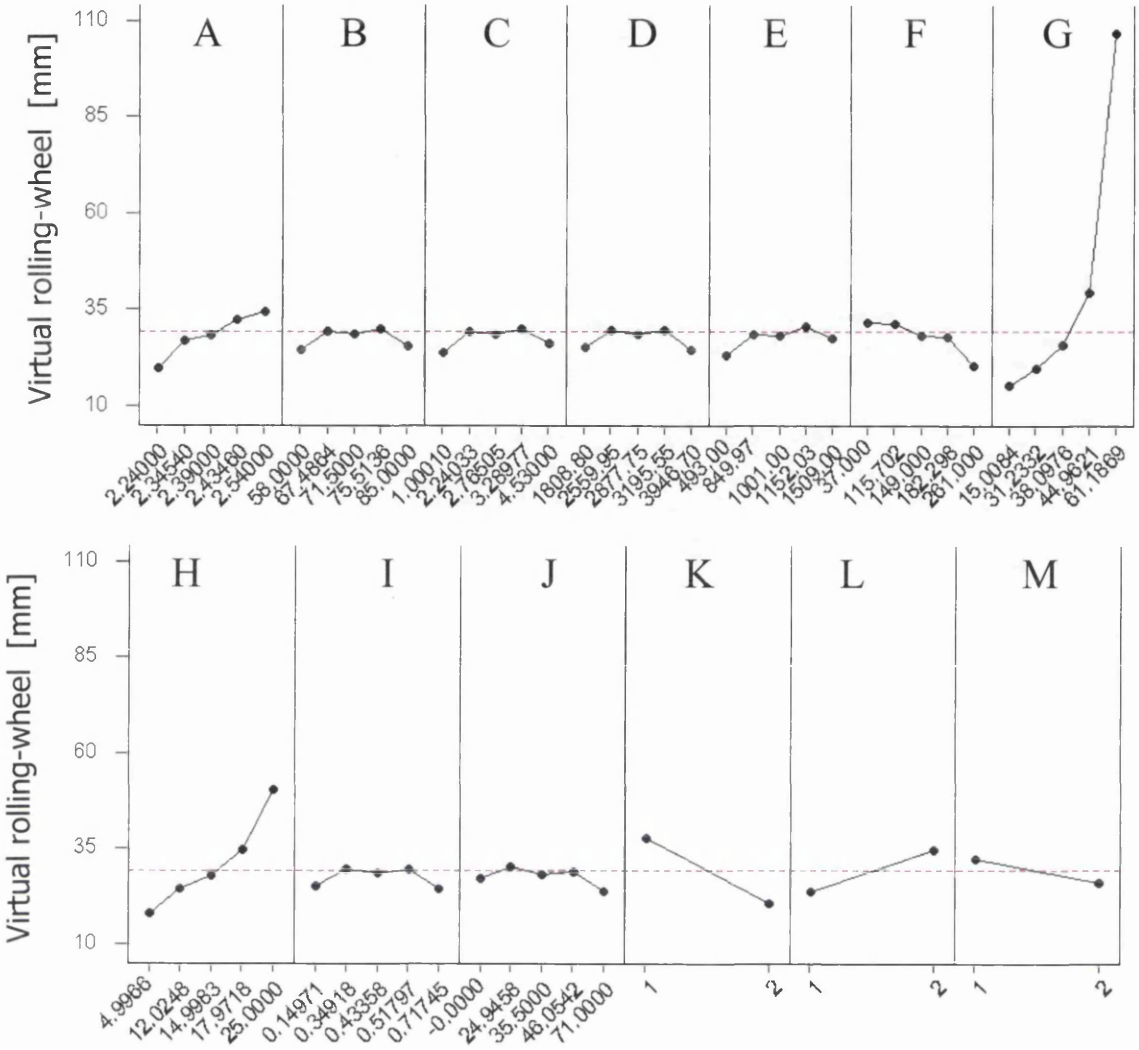
Both ANOVA's indicate strong and useful predictive models, which have many interacted terms. These DoE models were deemed to be useful for understanding the system.

To gain an overview of the effects of each factor, the main effects were calculated and plotted. The main effects are the response means for each factor level, in sorted order. A main effects plot shows these response means together with a horizontal line representing the overall experiment mean, the grand mean of all the response data. The effects are the differences between the response means and this line. These are shown for each response for adhesion in Figure 70 and for tack in Figure 71. They indicate the effect of moving through the factor levels on the responses. NB as the ANOVA's show that many of these factors are involved in interactions, the main effects plots do not show the influence of other factors on the one being studied, therefore some caution is required when analysing these plots.



Coded Factor Name	Factor Name	- level	+ level
A	Adhesive membrane thickness	2.24	2.54
B	Water Bath average Temp	58	85
C	UV Power 1st zone	1.0001	4.53
D	Total Radiation	1808.8	3946.7
E	Reaction Time	493	1509
F	Age of adhesive at time of coating	37	261
G	Resin A : Adhesive Polymer	15.008	61.18686
H	Resin B : Adhesive Polymer	4.9966	25
I	Resin C : Adhesive Polymer	0.1497	0.717447
J	Age of Tape at testing	0	71
K	Pilot-Scale or Full-scale Adhesive Polymer maker (1 or 2)	1	2
L	Pilot-Scale or Full-scale Compounding or Coating (1 or 2)	1	2
M	Who Tested, laboratory 1 or 2	1	2
N	Side of tape tested, Face or Liner (1 or 2)	1	2
O	Test, Adhesion or Rolling-Wheel tack (1 or 2)	1	2

Figure 70 Virtual adhesion main effects plot for each factor in the central composite DoE, together with a key.



Coded Factor Name	Factor Name	- level	+ level
A	Adhesive membrane thickness	2.24	2.54
B	Water Bath average Temp	58	85
C	UV Power 1st zone	1.0001	4.53
D	Total Radiation	1808.8	3946.7
E	Reaction Time	493	1509
F	Age of adhesive at time of coating	37	261
G	Resin A : Adhesive Polymer	15.008	61.18686
H	Resin B : Adhesive Polymer	4.9966	25
I	Resin C : Adhesive Polymer	0.1497	0.717447
J	Age of Tape at testing	0	71
K	Pilot-Scale or Full-scale Adhesive Polymer maker (1 or 2)	1	2
L	Pilot-Scale or Full-scale Compounding or Coating (1 or 2)	1	2
M	Who Tested, laboratory 1 or 2	1	2
N	Side of tape tested, Face or Liner (1 or 2)	1	2
O	Test, Adhesion or Rolling-Wheel tack (1 or 2)	1	2

Figure 71 Virtual rolling-wheel tack main effects plot for each factor in the central composite DoE, together with a key.

Visual inspection of the main effects plots reveals:-

1. Factors G and H have dramatic effects on both responses: increasing the amount of resins A and B increases the adhesion and the rolling-wheel tack. The larger the rolling-wheel number, the lower the tack. As resins A and B are called tackifying resins then unexpected behaviour is observed. Previously it was thought that increasing the amount of tackifier would make the tape tackier. The rolling wheel analysis reverses this assumption. Either the assumption that rolling-wheel measures levels of tack is wrong or more likely the assumptions about the tack modifying properties of these tackifiers is wrong.
2. The directions of effects slopes for each factor is the same for both responses, in line with the previously observed phenomenon that there is a strong degree of dependency between these two responses, i.e. an increase of adhesion is matched by an increase of rolling-wheel (reduction of tack).
3. The virtual product suffers from ageing effects. Factor F, age of adhesive at time of coating, shows a decrease in the levels of adhesion and rolling-wheel as the adhesive polymer ages. This is of significance because the adhesive polymer is made in batches, infrequently, and shipped trans-Atlantic. This method of supply means that there could be a variety of ages of adhesive polymer at the time of coating. Immunisation against this phenomenon should be considered as a future optimisation goal.
4. Factors K and L, 'Pilot-Scale or Full-scale Adhesive Polymer maker (1 or 2)' and 'Pilot-Scale or Full-scale Compounding or Coating (1 or 2)' have an influence on the properties. There appears to be an offset depending on which process is used. This is useful knowledge when scaling up future similar products from pilot to full-scale.
5. The two different laboratories used to perform the different measurements also has an influence on both responses. Factor M, 'Who Tested, laboratory 1 or 2', also seems to offset the values.

6. Factor A also seems to have a strong influence. Factor A, the 'Adhesive membrane thickness', could be a variable used in the quest for deriving super properties. Thinner membranes than those experienced by the experiment may yield higher tack properties (lower rolling-wheel), perhaps divorced from the adhesion/rolling-wheel dependency.
7. There appears to be little difference between face-side and liner-side adhesions. (Factor N). This is to be expected as both face and liner side adhesive coatings are identical.

The product designers had given a specification for adhesion and rolling wheel tack based on their analysis of the performance of the tape produced by the pilot processes. The full-scale processes were required to match those specifications. The target for adhesion was 25 N/25mm and for rolling-wheel tack the target was 25mm. The tolerances were based on pilot-scale process capability. Other product properties, such as coating thickness, that were explored as factors in the neural network model, also had to be maintained within certain parameters. Therefore the model was interrogated to obtain process parameters that would make product that would match the requirements of the product specification. This was done by making use of the mathematical functions within the DoE software that manipulate the quadratic models for each response by systematically adjusting the levels of the equation terms (process variables) until a match is found. This would be flagged as a recommended set of running conditions. Analysis of the response data shown in Figure 66 with targets superimposed is shown in Figure 72. The virtual product meeting these requirements was never achieved during the central composite DoE, therefore the neural network model was interrogated to find conditions that would yield virtual product with these targets.

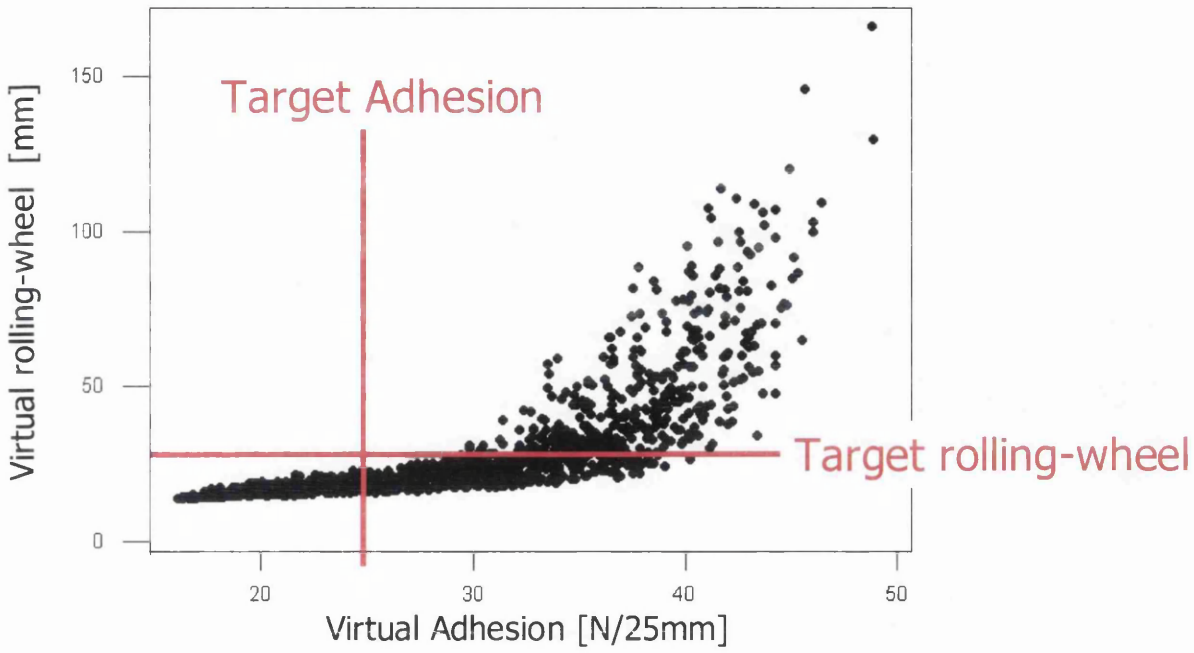


Figure 72 XY Plot of the virtual responses (liner-side only) for the central composite DoE data with target requirements superimposed.

The DoE quadratic equations in Table 23 and Table 24, for virtual adhesion and virtual rolling-wheel tack were examined to get solutions for the following criteria:-

1. Only use liner-side data
2. Only consider full-scale processes
3. Find solutions where both virtual responses are as close to their targets as possible
4. Fix factor F, Polymer age at time of coating, at 120 days, an expected typical value
5. Fix factor J, Days age after coating, at 70 days, roughly when the product is expected to be in use.

Face-Side Virtual Adhesion =

89.07184211	
-26.36683204	* A. Polymer packaging wall thickness
0.031877239	* B. Polymer reactor heating
-0.430240927	* C. UV power mW Model Zone 1
-0.00048829	* D. Total UV energy during polymer reaction
-0.000731659	* E. Polymer reaction time
0.003493416	* F. Polymer age at time of coating
-2.581939514	* G. Resin A:Polymer ratio
-3.296496354	* H, Resin B: polymer ratio
-0.049080726	* J. Days age after coating
0.012926025	* G. Resin A:Polymer ratio2
0.020997959	* H, Resin B: polymer ratio2
0.809455566	* A. Polymer packaging wall thickness * G. Resin A:Polymer ratio
0.942111448	* A. Polymer packaging wall thickness * H, Resin B: polymer ratio
0.03647091	* C. UV power mW Model Zone 1 * G. Resin A:Polymer ratio
-0.001282787	* F. Polymer age at time of coating * G. Resin A:Polymer ratio
0.034360195	* G. Resin A:Polymer ratio * H, Resin B: polymer ratio
0.001739665	* G. Resin A:Polymer ratio * J. Days age after coating

Liner-Side Virtual Adhesion =

96.82441295	
-26.36683204	* A. Polymer packaging wall thickness
0.031877239	* B. Polymer reactor heating
-0.430240927	* C. UV power mW Model Zone 1
-0.00048829	* D. Total UV energy during polymer reaction
-0.002226751	* E. Polymer reaction time
-0.00625013	* F. Polymer age at time of coating
-2.617078164	* G. Resin A:Polymer ratio
-3.374996784	* H, Resin B: polymer ratio
-0.028123299	* J. Days age after coating
0.012926025	* G. Resin A:Polymer ratio2
0.020997959	* H, Resin B: polymer ratio2
0.809455566	* A. Polymer packaging wall thickness * G. Resin A:Polymer ratio
0.942111448	* A. Polymer packaging wall thickness * H, Resin B: polymer ratio
0.03647091	* C. UV power mW Model Zone 1 * G. Resin A:Polymer ratio
-0.001282787	* F. Polymer age at time of coating * G. Resin A:Polymer ratio
0.034360195	* G. Resin A:Polymer ratio * H, Resin B: polymer ratio
0.001739665	* G. Resin A:Polymer ratio * J. Days age after coating

Table 23 DoE Equations predicting virtual adhesion

Liner-Side Virtual Rolling-Wheel Tack =	
820.0703655	
-266.0477146	* A. Polymer packaging wall thickness
-6.09350355	* C. UV power mW Model Zone 1
0.00428683	* E. Polymer reaction time
0.260061014	* F. Polymer age at time of coating
-21.24114132	* G. Resin A:Polymer ratio
-23.258511	* H, Resin B: polymer ratio
0.394182822	* J. Days age after coating
0.077083965	* G. Resin A:Polymer ratio2
0.113708085	* H, Resin B: polymer ratio2
5.872955669	* A. Polymer packaging wall thickness * G. Resin A:Polymer ratio
5.753564529	* A. Polymer packaging wall thickness * H, Resin B: polymer ratio
0.166104342	* C. UV power mW Model Zone 1 * G. Resin A:Polymer ratio
-0.00564506	* F. Polymer age at time of coating * G. Resin A:Polymer ratio
-0.006408698	* F. Polymer age at time of coating * H, Resin B: polymer ratio
0.212274751	* G. Resin A:Polymer ratio * H, Resin B: polymer ratio
-0.011509964	* G. Resin A:Polymer ratio * J. Days age after coating

Table 24 DoE Equations predicting virtual rolling-wheel

The results of the search are shown in the appendix. Any one of these 43 sets of conditions could yield responses at target values, with the given constraints.

At this stage the pragmatics of managing tight project timescales for new product launches, involving multinational resources and millions of pounds of investment meant that as soon as there was enough evidence that the full-scale coating and compounding process could produce consistent quality product, close to the target specification, then it was to be sold. The process settings had to be frozen at the then sub-optimum conditions as soon as the first commercial product was made. Product from the full-scale coating and compounding process was being sold before the second generation neural network model interrogation stage was complete. In a model validation exercise, these sub-optimum process conditions were replicated on the virtual process. A comparison of the virtual predictions of quality with the real measured quality would allow assessment of the model's predictive use. The sub-optimum process conditions are shown in Table 25. Factors F and J formed the axies of subsequent contour plots.

Factor	Value
A. Polymer packaging wall thickness	2.50
B. Polymer reactor heating	61.00
C. UV power mW Model Zone 1	3.50
D. Total UV energy during polymer reaction	2695.00
E. Polymer reaction time	770.00
F. Polymer age at time of coating	X Axis
G. Resin A:Polymer ratio	20.00
H, Resin B: polymer ratio	8.00
I. Additive:polymer ratio	0.28
J. Days age after coating	Y Axis
L. Adhesive Poly. maker. 1=Pilot 2=Full	2
M. Compounding & Coating 1=Pilot 2=Full	2
N. Lab testing 1 or 2 (production)	2
K. Side tested Face-Side=1 Liner-Side=2	2

Table 25 The sub-optimum levels of terms of the quadratic equations describing virtual adhesion and virtual rolling-wheel tack.

The first 56 test results, measured by the process operators, are shown in Figure 73 and Figure 74, in histogram format. They show that the adhesion mean is 25.7 and that the liner-side rolling-wheel mean is 27.7. These 56 test results represent tests made regularly over 206,000 m^2 of tape.

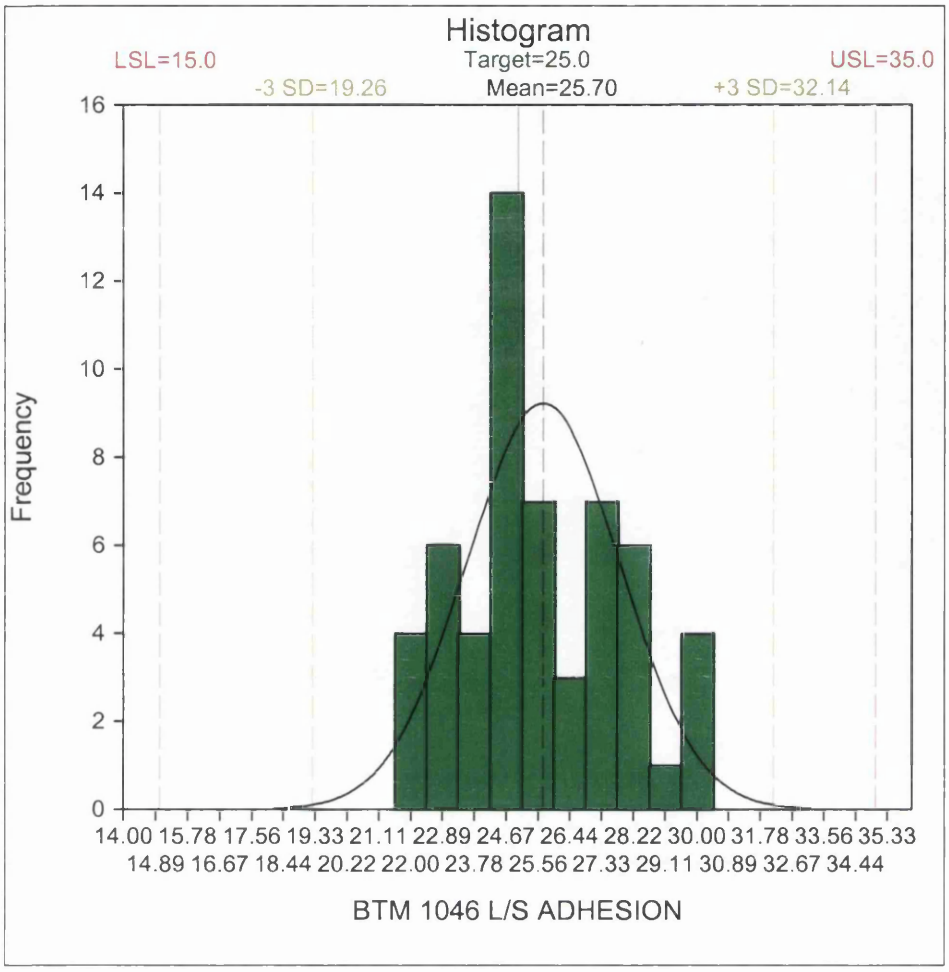


Figure 73 Histogram plot of the first 56 production test results for liner side adhesion.

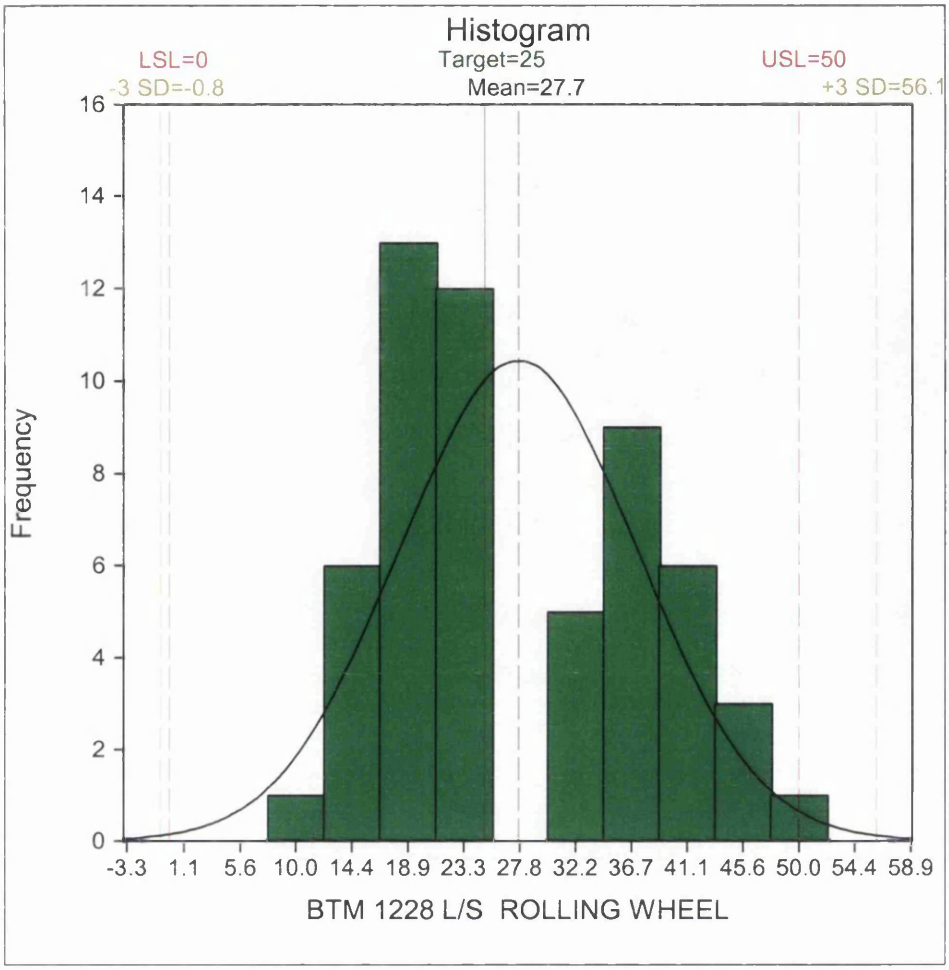


Figure 74 Histogram plot of the first 56 production test results for liner side rolling wheel.

The second generation neural network model predictions are close to these means. The prediction of virtual liner-side adhesion is about 26 for an adhesive polymer 120 days old, a typical value, and just under 30 for the rolling-wheel. The contour plots in Figure 75 and Figure 76 show how variation due to ageing effects would be expected from this process, depending on when the product was tested and how old the adhesive polymer was when it entered the compounding process. The agreement of the values validates the accuracy of the second generation neural network model under these conditions.

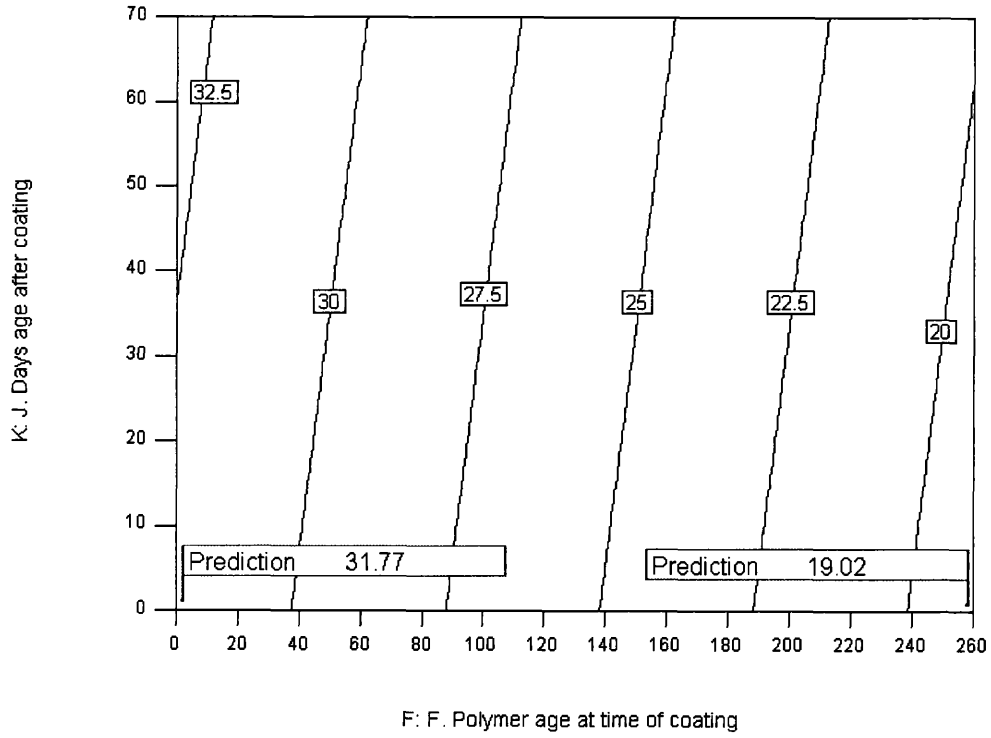


Figure 75 Contour plot of virtual liner-side adhesion showing predictions at the extremes of polymer age at the time of coating, tested fresh off the machine.

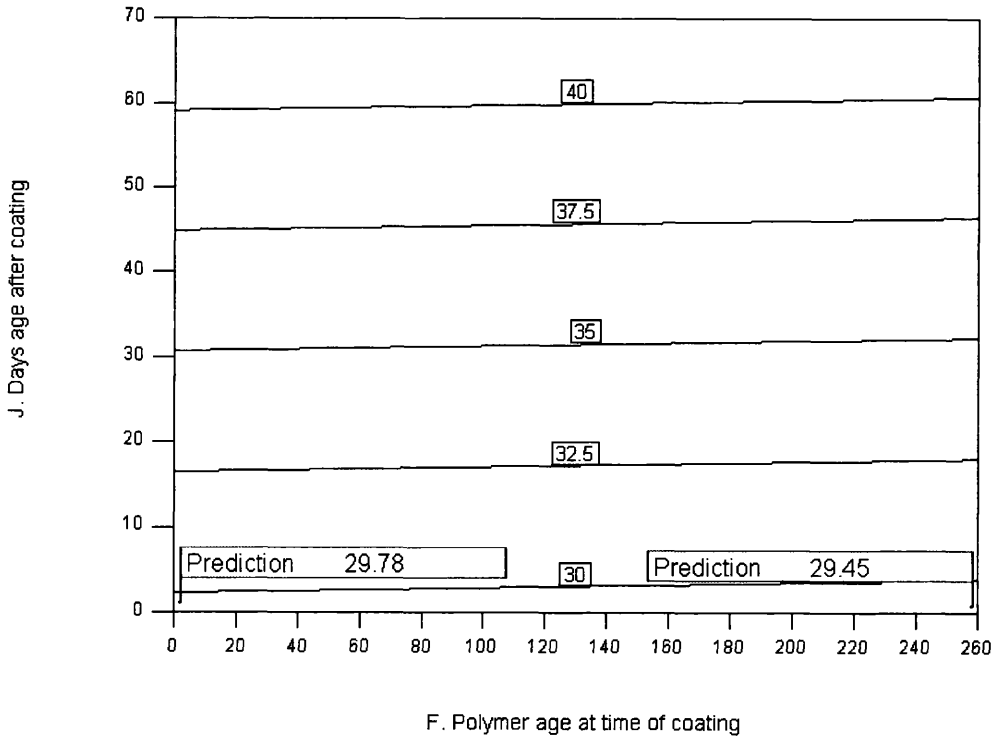


Figure 76Contour plot of virtual liner-side rolling-wheel tack showing predictions at the extremes of polymer age at the time of coating, tested fresh off the machine.

7.1.1.6.3 Potential 'super properties'

There is a high degree of dependence between adhesion and rolling-wheel tack , i.e. as adhesion increases the rolling-wheel increases (the actual tack decreases). 'Super tape' would completely break this dependency allowing high adhesion, low rolling-wheel and therefore high tack product. The DoE quadratic equations in Table 23 and Table 24, for virtual adhesion and virtual rolling-wheel tack were examined to get solutions for the following criteria describing super properties:-

1. Only use liner-side data
2. Only consider full-scale processes
3. Find solutions where both virtual responses are as large as possible
4. Fix factor F, Polymer age at time of coating, at 120 days, an expected typical value
5. Fix factor J, Days age after coating, at 70 days, roughly when the product is expected to be in use.

There are possible solutions. The best results of the search are shown in the table below.

	1	2	3
A. Polymer packaging wall thick	2.24	2.24	2.25
B. Polymer reactor heating	73.7	62.7	85
C. UV power mW Model Zone 1	4.53	3.82	3.98
D. Total UV energy during polym	3769.9	1846.82	3388.39
E. Polymer reaction time	504.22	563.25	759.02
F. Polymer age at time of coati	120.11	120.05	120
G. Resin A:Polymer ratio	61.18	61	61.18
H, Resin B: polymer ratio	9.91	7.59	9.37
I. Additive:polymer ratio*	0.4	0.16	0.52
J. Days age after coating	70	69.96	69.96
K. Side tested Face-Side=1 Line	2	2	2
Virtual Adhesion	50.1752	46.3526	48.7113
Virtual Rolling-Wheel Tack	36.085	24.1899	33.7512

Table 26 Conditions that yield optimum properties according to DoE analysis of the neural network model

Any one of these 3 sets of conditions could yield product with high adhesions and high tack results (low rolling-wheel) in regions well away from this dependent

behaviour. Their location in comparison with the data generated for the central composite DoE is shown in Figure 77 below where they are highlighted in red.

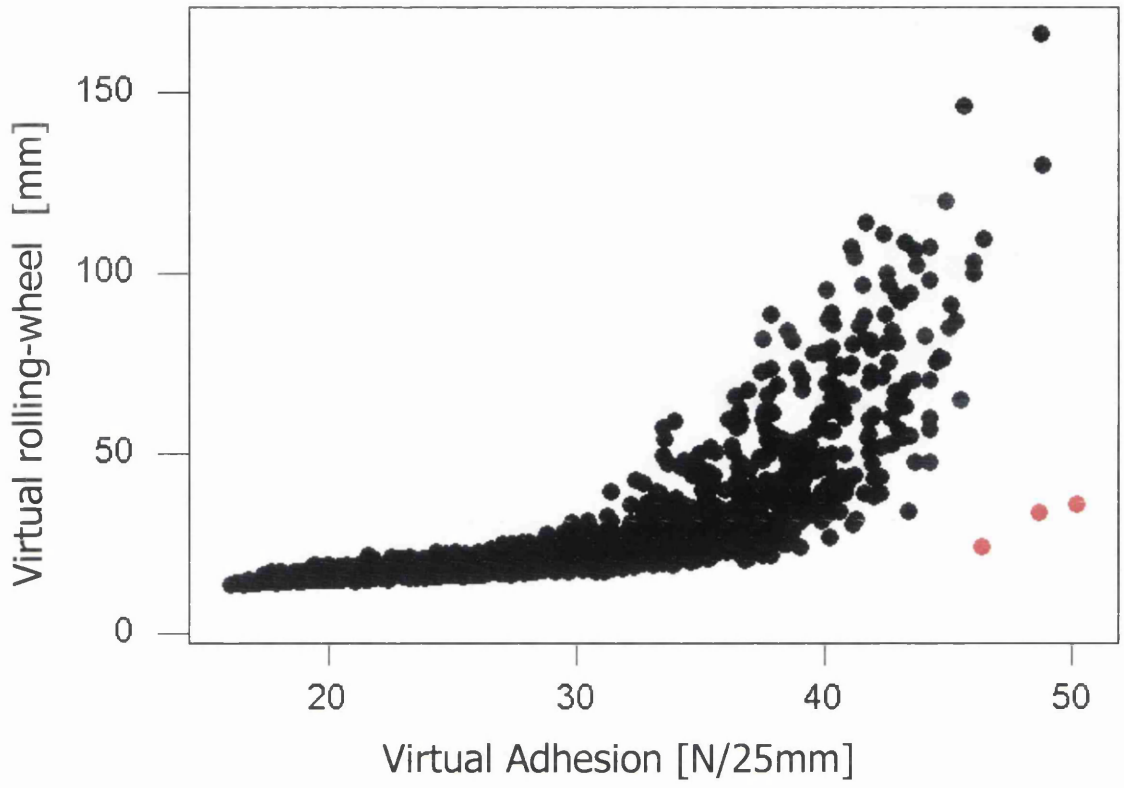


Figure 77 XY Plot of the virtual responses (face-side rolling-wheel data removed) for the central composite DoE data with the 3 super property predictions

7.2 Closure

It appears that the second generation neural network model is trustworthy and useful for predicting tape quality. Without the use of neural networks, fed with DoE structured data, the insights in to the contributions of both pilot and full-scale processes would have been limited to a few variables and their gross effects. The high level of technical understanding allowed management to be able to make calculated risks, knowing that the calculations were based on sound investigation. Two such decisions involved adjusting the formulation and reducing the coating thickness. The first was to improve thumb appeal, the second to reduce costs. Thumb appeal is apparent tack based on repeated touching with fingers; an unsophisticated test, full of

error, but one customers put great store in. Both decisions were checked virtually first for impact on properties.

Chapter 8 Discussion

The potential for the use of neural networks in the development of production processes has been demonstrated. The ability of neural networks to handle a large number of variables with relatively few data points gives it significant advantage compared with statistically based DoE techniques. Full factorial DoE's require 2^k runs to examine k factors. Fractional factorial DoE's, such as Taguchi orthogonal arrays, attempt to reduce the number of experimental runs required when k is large. However they make compromises with each degree of fractionation. For example a Taguchi L64 orthogonal array can determine main effects of 63 factors in 64 experimental runs, albeit at great risk of getting ambiguous results. In an L64, all main effects are confounded with many second order interactions. There is not enough data to resolve which factor or interaction is having the true effect. Taguchi orthogonal arrays also assume that terms add, usually linearly but sometimes as second order polynomials. By contrast, central composite DoE's are able to resolve all possible interactions, and make use of cubic equations to model with, but to match the Taguchi L64 for the low number of runs only 6 variables can be examined (albeit in great detail). With statistical techniques there is a relationship between the detail and integrity of the resulting model and the number of experiments required to derive that model. With neural networks, the relationship is similar in direction but more biased towards greater detail and integrity with less experimental runs. Neural networks can handle large numbers of factors, only limited by software constraints, without the data having to have the formal structure of DoE's. This makes them adept at combining small experiments and exploiting any historical data that would otherwise be unsuitable for DoE modelling. This feature gives them another powerful advantage. A neural network model can also adapt to changed circumstances by simply adding new data to the training data set and retraining the neural network. Because neural networks can handle all relevant factors, the effects of minor variables are not ignored, which often

they are when the magnified costs of running DoE experiments forces the experimenter not to peruse characterisation of minor contributing factors to a response. Neural networks therefore allow the development of models of complex relationships between parameters and output function.

The use of the experimental structure of DoE has enabled the training data sets to be gathered with the minimum number of real experiments yet still ensuring the data covered the majority of the operating envelope of the process. However, additional data points had to be generated to ensure the extreme conditions of production were covered. Traditional DoE techniques, such as orthogonal array, rely on the additive nature of the process to predict such extremes. However, the neural network can encompass all available data to enhance the level of prediction.

Full use was made of statistical tests to validate the quality of the model. These reduced the number of validation experiments required whilst providing confidence in the neural network model. They were also able to highlight where there were deficiencies in the quality of the input data, i.e. the face-side tack, due to anomalies in production. This prevented the model being based on data which had biases due to unknown factors.

The experience gained from the model of the pilot processes was used to shape the development of the full-scale processes. The modelling of the sub-processes enabled the model to be developed in parallel with the physical development of the process.

Although the neural network model was shown to be statistically valid, in order to obtain useful information for production optimisation and control, a series of structured experiments using DoE techniques had to be performed on the model, the virtual process. These generated operating envelopes and established the sensitivity of the process. As these were based on a virtual process, there was potential to run many

combinations at minimal cost. The use of structured DoE's on a virtual model was shown to be an efficient method for ensuring the product produced matched the specification. The ability to explore the full operating envelope opened the possibility for a product with enhanced "super" properties and also to make the process robust to variability in input raw materials or process controls.

Chapter 9 Conclusions and Recommendations for Future Work

9.1 Conclusions

Neural network modelling techniques have been successfully applied to minimise the development time for a full-scale production process. Neural network models of the pilot and production facilities have been created and validated. The following conclusions can be drawn:-

- Neural networks offer a flexible approach to deal with processes in which there are a large number of parameters, without the need for enormous numbers of experimental runs or the need to simplify the model.
- The structured approach of DoE techniques was used to enable efficient generation of data that covered the full operating envelope. However additional data points were necessary to ensure all extremes were covered.
- Once the neural network model was created, traditional DoE experiments had to be run on the virtual process to establish optimum operating conditions, quantify the effect of parameters and response surfaces for process control.
- The neural network model predicted a potential "super" tape with enhanced properties.

9.2 Future Work

- Only 10 numeric variables were used to analyse the second generation neural network model. The neural network model knows about 33 variables in total. There may be a lot more insight to be gained by interrogating the neural network using DoE's made up of different factors and variable levels.

- The potential for a 'super property' product should be explored to validate the neural network model.

- There seemed to be a lot of random noise from the system. An investigation in to tightening the test methods for adhesion and rolling-wheel tack should be considered in order to improve the signal to noise ratio.

- The mystery variable affecting face-side rolling-wheel tack was never identified. A possible clue could be in the fact that only the face side of the tape was affected. The face side is coated last and embossed with the surface patterns of various face rollers before it is wound. The first face roller that the adhesive contacts is a chill roll that will freeze any surface morphology. Also small coating patterns on the adhesive could cause the wheel to have less contact, therefore less stopping power. These coating patterns are caused by the angle of the die, the penetration of the die and the viscosity of the adhesive. The liner side would not suffer from surface irregularities as the liner-side surface is coated on to a very smooth liner; the only embossing possible is from the liner. Therefore the rolling-wheel experiences a glass-like adhesive, maximising the stopping effect.

- The ageing effects were never explained or minimised.

- The neural network model indicates that depending which laboratory did the testing, generally higher or lower results are obtained. Various statistical tools exist to aid investigation in to the causes of offset between laboratory results. If changes are to be made to the measurement system, it is vital that at least one of the laboratory's procedures remain unchanged until a future version of the second generation neural network model can be 'informed' of the changes if the model is going to be of future use.

Bibliography

D.J. Wheeler & D.S. Chambers, *Understanding Statistical Process Control 2nd Edition,* SPC Press, USA, 1992

ERA Technology Seminar, *Introduction to Applying Neural Networks, 1997*

NIST SEMATECH, *Engineering Statistics Handbook*,
URL<http://www.itl.nist.gov/div898/handbook/>

StatSoft, Inc. (1997). Electronic Statistics Textbook. Tulsa, OK: StatSoft.
URL<http://w ww.statsoft.com/textbook/stathome.html>

References

[1] J.G. Taylor, *The promise of Neural Networks,* London, Springer-Verlag, 1993.

[2] Bishop, C.M. *Neural Networks for Pattern Recognition,* Oxford University Press, 1995.

[3] L. Tarassenko , *A Guide to Neural Computing Applications,* John Wiley & Sons, 1998

[4] Fogleman and Gallinari, *Industrial Applications of Neural Networks,* Singapore, World Scientific, 1998

[5] Raj et al, *Modelling of Many Processes with ANNs for Intelligent Manufacture,* Int Jnl of Machine Manufacture Vol 40, Issue 6, May 2000

[6] Moser et al, *Modeling the Functional performance of Plasma Polymerized thin Films*, Thin Solid Films, Volumes 255-256, 1999

[7] Yu-To Chen et al, *A Neural Network Predictor For Reengineering Of Resin Manufacturing*, <URL:http://citeseer.nj.nec.com/cache/papers/cs/18498/http:zSzzSztardis.union.eduzSz~cheny2zSzPdf zSzpredictor3.pdf/a-neural-network-predictor.pdfCiteSeer>,[Accessed 13/12/02]

[8] Linda Trocine and Linda C. Malone, 2000, Finding Important Independent Variables Through Screening Designs: a Comparison of Methods, [WWW] <URL: http://www.informs-s.org/wsc00papers/098.PDF>, [accessed 13/12/02]

[9] Vaidyanathan et al, 2000, *Neural Network and Response surface Methodology for Rocket Engine Component Optimization*, 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Paper No. AIAA 2000-4880, 2000

[10] Nandi et al, *Artificial Neural-Network-Assisted Stochastic Process Optimization Strategies*, AIChE Journal Vol.47, January 2001

[11] Thomas Pyzdek, March 1999, Virtual-DOE, Data Mining and Artificial Neural Networks, QUALITYAMERICA.COM,

<URL:http://www.qualityamerica.com/knowledgecente/articles/PYZDEKneural.htm> [Accessed 13/12/02]

[12] Chu et al, 2002, *Constrained Optimization of Combustion in Simulated Coal-Fired Boiler Using Artificial Neural Network Models and Iformation Analysis*, Elsevier Science,<URL: http://www.sciencedirect.com/science?_ob=ArticleURL&_aset=A-WA-A-A-MsSAYWW-UUA-AUVUCCEADD-CAYVAUVYD-A-U&_origAset=A-WA-A-A-MsSAYWW-UUA-AUVUCCYDYU-CAYVUZEUV-A-U&_rdoc=1&_fmt=full&_udi=B6V3B-4772N0J-3&_coverDate=12%2F13%2F2002&_cdi=5726&_orig=search&_st=13&_sort=d&_acct=C000000453&_version=1&_urlVersion=0&_userid=3132&md5=189ac2e88765c8dd1710ccf9f1655122>

[13] Del Vecchio, *Understanding Design of Experiments*, Hanser/Gardner publication, Inc, USA 1997

[14] Donald J. Wheeler, *Understanding Industrial experimentation Second Edition*, SPC Press Inc, USA, 1990

[15] NeuralWare, *Professional II/Plus Software Manual*, NeuralWare Inc. USA, 1996

[16] Minitab Statistical Software Release 13.30, *Help File*

[17] Clarke J. Hall Jnr, Private communication, 3M, 1996-7

Appendices

Appendix 1 Conditions that give virtual product at their targets, derived form the interrogating central composite DoE run on the second generation neural network model.

Number	A. Polymer packaging wall thickness	B. Polymer reactor heating	C. UV power mW Model Zone 1	D. Total UV energy during polymer reaction	E. Polymer reaction time	F. Polymer age at time of coating *	G. Resin A:Polymer ratio	H, Resin B: polymer ratio	I. Additive:polymer ratio*	J. Days age after coating *	K. Side tested Face-Side=1 Liner-Side=2 *	Virtual Adhesion	Virtual Rolling-Wheel Tack
1	2.39	80.76	4.00	2390.00	725.83	120.00	31.73	7.24	0.71	70.00	2	25	25
2	2.30	59.06	4.53	2020.33	1502.91	120.00	37.75	5.00	0.51	70.00	2	25	25.0004
3	2.30	81.62	3.55	2021.09	1221.74	120.00	36.89	5.15	0.51	69.99	2	25	25.0002
4	2.47	63.75	1.01	3894.12	1492.29	120.00	28.05	19.91	0.41	70.00	2	25	25.025
5	2.41	63.68	1.01	3153.51	1505.65	120.00	25.57	23.46	0.29	69.94	2	25	25
6	2.32	81.37	3.82	2723.25	1172.83	120.00	36.53	5.00	0.45	70.00	2	24.9995	25.155
7	2.49	78.00	1.03	3944.27	1508.01	121.39	24.86	22.42	0.60	70.00	2	25	25.0716
8	2.44	80.90	4.46	3645.61	977.97	120.00	33.52	5.04	0.72	69.95	2	24.7949	24.9997
9	2.46	68.50	1.22	3946.64	1508.57	120.00	30.41	18.62	0.24	69.93	2	26.006	24.9999
10	2.50	63.97	1.00	2340.50	1498.18	120.00	27.53	20.28	0.49	70.00	2	26.0098	25.3669
11	2.42	73.45	1.20	3690.14	1509.00	121.02	29.69	20.05	0.15	70.00	2	26.1472	24.8618
12	2.42	62.97	4.53	3502.63	767.42	120.00	32.96	5.56	0.47	70.00	2	24.5237	25
13	2.54	58.00	1.11	3609.87	1180.61	119.94	23.33	24.97	0.36	69.64	2	26.8446	25
14	2.36	80.02	4.10	1914.63	1506.36	120.00	36.68	5.07	0.30	62.79	2	25	25.6135
15	2.43	83.81	1.00	3153.70	1269.52	120.95	24.67	25.00	0.15	69.14	2	27.0112	24.9466
16	2.32	84.80	2.81	3354.65	881.68	120.00	36.01	5.00	0.62	70.00	2	24.2891	25
17	2.38	84.85	3.08	3647.20	854.83	120.26	34.55	5.59	0.15	69.79	2	24.3013	25
18	2.54	61.16	1.00	1809.16	1314.61	121.16	30.89	17.51	0.71	70.00	2	27.7581	24.9967
19	2.40	77.89	1.19	2704.50	1509.00	120.00	31.75	19.45	0.38	70.00	2	27.5108	24.7823
20	2.42	77.34	2.22	3282.25	1507.55	119.99	26.08	25.00	0.46	70.00	2	27.9519	24.9943
21	2.31	58.10	4.02	2920.91	1355.30	119.96	37.17	5.07	0.58	70.00	2	24.1706	25
22	2.54	84.64	1.00	3437.63	1155.12	120.11	28.93	19.64	0.20	69.71	2	28.185	25
23	2.53	58.00	1.01	1811.14	1490.28	120.00	36.43	13.31	0.56	70.00	2	28.4589	25
24	2.53	79.36	1.00	2734.77	1008.34	120.00	23.85	25.00	0.40	69.99	2	28.56	25
25	2.54	59.03	2.01	2806.70	1241.70	120.00	28.38	20.89	0.15	70.00	2	28.6896	25.0026
26	2.53	75.15	1.87	3016.98	1168.23	120.00	24.58	25.00	0.69	70.00	2	29.0827	25
27	2.54	77.94	3.23	3946.50	1466.39	118.54	29.43	20.24	0.51	69.96	2	29.4501	25
28	2.45	71.68	1.93	3946.69	1508.99	120.00	31.28	15.66	0.67	69.97	2	25.0001	22.8415
29	2.47	84.88	3.35	3946.70	1501.19	120.00	34.25	16.83	0.39	69.82	2	30.1063	25
30	2.48	79.83	1.00	2247.35	1508.98	120.00	39.95	12.59	0.60	70.00	2	30.323	25
31	2.39	58.00	3.47	2080.41	1490.01	120.00	31.56	21.47	0.15	68.96	2	30.3189	24.7897
32	2.48	59.89	4.22	2717.09	1508.98	120.00	36.96	14.11	0.64	70.00	2	31.1312	25
33	2.54	69.82	1.38	2230.65	603.83	120.03	25.35	25.00	0.50	70.00	2	31.2266	25
34	2.28	84.64	2.35	3717.15	1050.92	114.12	36.83	5.00	0.18	70.00	2	23.6046	25
35	2.41	62.14	1.74	2172.76	672.85	120.00	28.71	25.00	0.64	70.00	2	31.7702	25
36	2.36	60.23	1.87	1812.21	1254.95	118.61	40.02	16.42	0.25	69.30	2	32.6113	25.0001
37	2.37	85.00	1.11	2320.80	1505.16	120.90	43.01	13.98	0.15	63.44	2	32.0089	25
38	2.48	84.63	4.50	3945.00	1508.99	120.07	42.95	7.99	0.22	61.42	2	31.9581	24.987
39	2.26	82.07	4.46	2356.16	1509.00	120.00	33.62	24.78	0.72	69.89	2	33.9723	25
40	2.49	58.01	1.92	3434.35	739.13	119.99	32.67	5.02	0.16	70.00	2	22.5247	24.9998
41	2.54	83.09	1.45	1808.80	1508.95	123.52	43.58	8.00	0.23	53.74	2	30.3897	24.9997
42	2.27	84.29	2.86	1974.09	1150.14	120.00	43.87	17.23	0.57	69.98	2	36.9719	25
43	2.54	70.64	4.19	2160.13	534.18	120.00	42.78	10.11	0.28	70.00	2	37.2359	25

Appendix 2 Details of experiments used to develop the first generation neural network model

The data matrix shown below is provided to help illustrate how the DoE's relate to each other. Each row of the matrix represents different conditions under which product was made and tested for adhesion and tack. Each column in the matrix represents a non-correlating, adequately varied process variable suitable for generating the neural network model with. Variables unsuited for neural network modelling have been omitted. The columns are grouped by process, i.e. the variables grouped with blue in the first row are from the adhesive polymer maker and those grouped with red are from the compounding and coating processes.

	ADHESIVE POLYMER MAKER												COATING & COMPOUNDING									
Variable description	Side of tape tested 1/s=1 l/s=2	Gas Flow [Vol per unit time]	Mass of A used in formulation	Antioxidant pph of total monomers	B C used in formulation	A C used in formulation	D C used in formulation	Heating °	Agitation	Radiation Received	Reaction time	Average radiation source intensity	aerobic/ anearobic 1/2	Minibatch size	Minibatch mass	adhesive age at time of coating	Coater uv Dose	tackifier1 : adhesive polymer	tackifier2 : adhesive polymer	additive adhesive polymer	Extruder rpm / kg/hr	Time under UV s	average total thickness of coating
	1	0	8	0.5	100	17.78	4.44	85	0	2578	810	3.183	1	79	366	129	1000	35	15	0.5	27.2	1.836735	313
	1	0	8	0.5	100	17.78	4.44	85	0	2563	567	4.521	1	79	367.7	121	1000	35	15	0.5	26.6	1.8	313
DoE 2^6-1 ADHESIVE POLYMER MAKER Formulation and process condition experiment	2	20	6.7	0.3	64.99	12.03	2.51	74	80	2365	671	3.52459	2	77.5	314.5	143	1000	35	15	0.5	30.8	1.8	278
	2	20	4.8	0.3	134.91	11.97	2.49	73	80	2786	786	3.544529	2	77.375	303.15	143	1000	35	15	0.5	28.6	1.836735	300
	2	20	12.5	0.3	65.13	23.95	2.49	73	80	2786	786	3.544529	2	78.25	304.3	142	1000	35	15	0.5	28.8	1.8	293
	2	20	9.2	0.3	134.73	24.02	2.35	73	80	2365	671	3.52459	2	80	309.85	143	1000	35	15	0.5	30.3	1.764706	277
	2	20	6.5	0.3	64.95	11.93	6.42	73	80	2773	786	3.527354	2	79	309.2	138	1000	35	15	0.5	28.3	1.8	298
	2	20	4.7	0.3	134.94	11.9	6.33	73	80	2365	671	3.52459	2	80	307.15	138	1000	35	15	0.5	29.7	1.836735	291
	2	20	12.3	0.3	65.04	24.02	6.45	73	80	2365	671	3.52459	2	80	313.1	137	1000	35	15	0.5	35.9	1.8	245
	2	20	9	0.3	135.01	23.87	6.37	73	80	2773	786	3.527354	2	39.5	296.6	137	1000	35	15	0.5	41.3	1.8	220
	2	20	6.7	0.7	64.99	12.03	2.51	73	80	2773	786	3.527354	2	79.5	305.75	137	1000	35	15	0.5	28.1	1.8	299
	2	20	4.8	0.7	134.91	11.97	2.49	73	80	2357	671	3.512668	2	79.5	309.05	137	1000	35	15	0.5	27.6	1.836735	309
	2	20	12.5	0.7	65.13	23.95	2.49	74	80	2357	671	3.512668	2	79.5	313.9	137	1000	35	15	0.5	30.6	1.8	279
	2	20	9.2	0.7	134.73	24.02	2.35	73	80	2773	786	3.527354	2	79.25	301.5	136	1000	35	15	0.5	32.1	1.764706	264
	2	20	6.5	0.7	64.95	11.93	6.42	74	80	2357	671	3.512668	2	79.25	315.45	136	1000	35	15	0.5	27.6	1.8	304
	2	20	4.7	0.7	134.94	11.9	6.33	73	80	2773	786	3.527354	2	79.25	301.75	136	1000	35	15	0.5	36.5	1.836735	246
	2	20	12.3	0.7	65.04	24.02	6.45	73	80	2773	786	3.527354	2	80	303.25	136	1000	35	15	0.5	30.9	1.8	277
	2	20	9	0.7	135.01	23.87	6.37	73	80	2357	671	3.512668	2	79.75	314	136	1000	35	15	0.5	29.4	1.875	298
	2	20	6.7	0.3	64.99	12.03	2.51	73	80	2773	786	3.527354	2	80	425.8	135	1000	35	15	0.5	36.4	1.8	243
	2	20	4.8	0.3	134.91	11.97	2.49	74	80	2357	671	3.512668	2	80	426.45	135	1000	35	15	0.5	40.8	1.8	222
	2	20	12.5	0.3	65.13	23.95	2.49	74	80	2357	671	3.512668	2	79.75	428.45	135	1000	35	15	0.5	29.2	1.875	300
	2	20	9.2	0.3	134.73	24.02	2.35	73	80	2785	786	3.543257	2	79.75	432.1	134	1000	35	15	0.5	31.7	1.8	271
	2	20	6.5	0.3	64.95	11.93	6.42	74	80	2357	671	3.512668	2	80	434.75	134	1000	35	15	0.5	30.4	1.8	281
	2	20	4.7	0.3	134.94	11.9	6.33	73	80	2785	786	3.543257	2	78.5	328.6	134	1000	35	15	0.5	30.9	1.875	286
	2	20	12.3	0.3	65.04	24.02	6.45	73	80	2785	786	3.543257	2	78.5	336.1	134	1000	35	15	0.5	28.8	1.836735	298
	2	20	9	0.3	135.01	23.87	6.37	74	80	2357	671	3.512668	2	79	433.65	133	1000	35	15	0.5	30.4	1.8	281
	2	20	6.7	0.7	64.99	12.03	2.51	74	80	2357	671	3.512668	2	80	432.15	133	1000	35	15	0.5	28.4	1.836735	302
	2	20	4.8	0.7	74.12	8.87	1.85	73	80	2785	786	3.543257	2	78.75	432.85	133	1000	35	15	0.5	28.5	1.8	296
	2	20	12.5	0.7	65.13	23.95	2.49	74	80	2785	786	3.543257	2	78.125	435.4	133	1000	35	15	0.5	31.9	1.8	270
	2	20	9.2	0.7	134.73	24.02	2.35	74	80	2357	671	3.512668	2	79	430.9	133	1000	35	15	0.5	30.4	1.8	281
	2	20	6.5	0.7	64.95	11.93	6.42	74	80	2785	786	3.543257	2	78.25	432.15	133	1000	35	15	0.5	29.8	1.875	295
	2	20	4.7	0.7	134.94	11.9	6.33	74	80	2357	671	3.512668	2	78.5	430.95	133	1000	35	15	0.5	28.2	1.836735	304
	2	20	12.3	0.7	65.04	24.02	6.45	73	80	2785	786	3.543257	2	79.25	439.3	132	1000	35	15	0.5	28.9	1.836735	298
	2	20	9	0.7	135.01	23.87	6.37	74	80	2357	671	3.512668	2	80	433.4	132	1000	35	15	0.5	33.9	1.8	257
DoE 2^6-1 ADHESIVE POLYMER MAKER Formulation and process condition experiment	2	0	8	0.5	100	17.78	4.44	62	0	2570	864	2.975	1	78	364.5	130	1000	35	15	0.5	30.2	1.8	282
	2	40	8	0.5	100	17.78	4.44	61	161	2570	864	2.975	2	77.75	370.55	130	1000	35	15	0.5	28.8	1.875	303
	2	40	8	0.5	100	17.78	4.44	85	0	2588	864	2.995	2	79.5	370.95	129	1000	35	15	0.5	32.1	1.875	278
	2	0	8	0.5	100	17.78	4.44	85	161	2588	864	2.995	1	80	373.25	130	1000	35	15	0.5	29.7	1.836735	291
	2	40	8	0.5	100	17.78	4.44	61	0	2578	810	3.183	2	78.5	371.2	129	1000	35	15	0.5	32.1	1.836735	273
	2	0	8	0.5	100	17.78	4.44	60	161	2578	810	3.183	1	78.5	363	128	1000	35	15	0.5	38.1	1.8	234
	2	0	8	0.5	100	17.78	4.44	85	0	2578	810	3.183	1	79	366	129	1000	35	15	0.5	28.9	1.836735	297
	2	40	8	0.5	100	17.78	4.44	85	161	2578	810	3.183	2	79.75	368.35	129	1000	35	15	0.5	28.1	1.8	300
	2	40	8	0.5	100	17.78	4.44	61	0	2591	895	2.895	2	78.25	368.15	128	1000	35	15	0.5	30.6	1.8	279
	2	0	8	0.5	100	17.78	4.44	61	161	2591	895	2.895	1	77.625	370.2	128	1000	35	15	0.5	29.9	1.8	284
	2	0	8	0.5	100	17.78	4.44	85	0	2591	895	2.895	1	79.375	369.25	127	1000	35	15	0.5	30.7	1.836735	283
	2	40	8	0.5	100	17.78	4.44	85	161	2591	895	2.895	2	78	368.5	127	1000	35	15	0.5	29.7	1.875	296
	2	0	8	0.5	100	17.78	4.44	61	0	2570	739	3.477	1	80	364	126	1000	35	15	0.5	28.8	1.836735	298
	2	40	8	0.5	100	17.78	4.44	61	161	2570	739	3.477	2	79	364	127	1000	35	15	0.5	28.4	1.836735	302
	2	40	8	0.5	100	17.78	4.44	85	0	2570	739	3.477	2	79.75	366.2	127	1000	35	15	0.5	28.4	1.8	297
	2	0	8	0.5	100	17.78	4.44	85	161	2570	739	3.477	1	79.5	368	127	1000	35	15	0.5	28.3	1.836735	303
	2	40	8	0.5	100	17.78	4.44	61	0	2319	593	3.911	2	80.5	360.5	126	1000	35	15	0.5	28.5	1.875	306
	2	0	8	0.5	100	17.78	4.44	61	161	2319	593	3.911	1	79.75	368.05	126	1000	35	15	0.5	30.9	1.836735	281
	2	0	8	0.5	100	17.78	4.44	85	0	2319	593	3.911	1	79.5	360.25	126	1000	35	15	0.5	30.2	1.836735	287
	2	40	8	0.5	100	17.78	4.44	85	161	2319	593	3.911	2	80	364.2	126	1000	35	15	0.5	38.1	1.836735	238
	2	0	8	0.5	100	17.78	4.44	61	0	2548	567	4.493474	1	80	363	125	1000	35	15	0.5	29.2	1.836735	295
	2	40	8	0.5	100	17.78	4.44	61	161	2548	567	4.493474	2	80	365.5	125	1000	35	15	0.5	28.9	1.836735	297
	2	40	8	0.5	100	17.78	4.44	85	0	2548	567	4.493474	2	81	362	125	1000	35	15	0.5	30.6	1.836735	284
	2	0	8	0.5	100	17.78	4.44	85	161	2548	567	4.493474	1	80	368.3	125	1000	35	15	0.5	28.7	1.875	305
	2	0	8	0.5	100	17.78	4.44	61	0	2572	654	3.932	1	80	371.9	124	1000	35	15	0.5	33.4	1.875	269
	2	40	8	0.5	100	17.78	4.44	60	161	2572	654	3.932	2	79.75	369.75	124	1000	35	15	0.5	33.1	1.8	262
	2	40	8	0.5	100	17.78	4.44	85	0	2572	654	3.932	2	79.5	370.65	124	1000	35	15	0.5	29.2	1.836735	295
	2	0	8	0.5	100	17.78	4.44	85	161	2572	654	3.932	1	79.5	370.5	124	1000	35	15	0.5	29.4	1.875	299
	2	40	8	0.5	100	17.78	4.44	62	0	2563	567	4.521	2	78.5	367	121	1000	35	15	0.5	26.9	1.8	311
	2	0	8	0.5	100	17.78	4.44	61	161	2563	567	4.521	1	78.5	363	121	1000	35	15	0.5	28.8	1.8	293
	2	0	8	0.5	100	17.78	4.44	85	0	2563	567	4.521	1	79	367.7	121	1000	35	15	0.5	29.9	1.8	284
	2	40	8	0.5	100	17.78	4.44	85	161	2563	567	4.521	2	78.5	363	121	1000	35	15	0.5	31.9	1.875	279
Doe 2^5-1 COATING & COMPOUNDING Formulation and process condition	2	20	8	0.5	100	17.78	4.44	73	80	2561	729	3.513443	2	80	365	120	1000	25	5	0.3	31.9	1.5	233
	2	20	8	0.5	100	17.78	4.44	74	80	2561	729	3.513443	2	79	365.6	120	800	45	5	0.3	28.9	1.5	252
	2	20	8	0.5	100	17.78	4.44	73	80	2561	729	3.513443	2	79	367.5	120	800	25	25	0.3	29.6	1.5	247
	2	20	8	0.5	100	17.78	4.44	73	80	2561	729	3.513443	2	79	366.8	120	1000	45	25	0.3	28	1.47541	271
	2	20	8	0.5	100	17.78	4.44	73	80	2561	729	3.513443	2	78.5	369.1	120	800	25	5	0.7	34.2	1.525424	224
	2	20	8	0.5	100	17.78	4.44	74	80	2561	729	3.513443	2	78.25	370.2	120	1000	45	5	0.7	28.4	1.5	256
	2	20	8	0.5	100	17.78	4.44	74	80	2561	729	3.513443	2	80	370.5	119	1000	25	25	0.7	31.4	1.5	236
	2	20	8	0.5	100	17.78	4.44	73	80	2561	729	3.513443	2	79	369.5	119	750	45	25	0.7	25.9	1.45161.3	268
	2	20	8	0.5	100	17.78	4.44	74	80	2561	729	3.513443	2	79	369.25	119	600	25	5	0.3	32.2	2.25	322
	2	20	8	0.5	100	17.78	4.44	73	80	2561	729	3.513443	2	78.5	370.85	119	1000	45	5	0.3	30.4	2.25	338
	2	20	8	0.5	100	17.78	4.44	73	80	2561	729	3.513443	2	78.75	368	119	1000	25	25	0.3	29.9	2.25	343
	2	20	8	0.5	100	17.78	4.44	74	80	2561	729	3.513443	2	78.5	367.5	119	600	45	25	0.3	26.1	2	348
	2	20	8	0.5	100	17.78	4.44	74	80	2561	729	3.513443	2	80	368.5	119	1000	25	5	0.7	32.2	2.307692	329
	2	20	8	0.5	100	17.78	4.44	74	80	2561	729	3.513443	2	79.75	366	119	750	45	5	0.7	27.4	2.25	369
	2	20	8	0.5	100	17.78	4.44	74	80	2561	729	3.513443	2	79.25	370.15	119	750	25	25	0.7	30.7	2.25	336
	2	20	8	0.5	100	17.78	4.44	74	80	2561	729	3.513443	2	79	370	119	950	45	25	0.7	28	2	328

Each DoE data set is grouped by first column colour. For example the first pilot-scale adhesive polymer maker DoE has rows that are grouped together by a pale yellow first column. Within each DoE data set, the coloured columns represent the experimental factors and their high and low levels. The uncoloured columns represent

the other process variables that were either held constant during individual DoE's and altered between, or allowed to drift. The ones allowed to drift, such as mini batch size, were the consequence of other sub-processes but considered as suitable summaries of those sub processes. The responses of those sub-processes were considered as suitable variables for this study.

Strict adherence to the DoE methodology of having only prescribed high and low level values within an experiment was not always possible. This is handled when analysis is done with a neural network.