A statistical model of an ammonium nitrate fluidised bed granulator

RL Kok

22793836

Dissertation submitted in fulfilment of the requirements for the degree Masters in Chemical Engineering at the Potchefstroom Campus of the North-West University

Supervisor: Mr AF van der Merwe

Co-supervisor: Prof G van Schoor

November 2016

It all starts here ™

NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT

Acknowledgements

The Lord God Almighty for strength, knowledge, patience and endurance

Omnia Holdings Limited, Sasolburg, South Africa

- Experimental setup and financial support
- Production team for their assistance
- Quality control lab for analytical equipment
- Mr Imtiaz Laher for guidance and assistance

North-West University, Potchefstroom, South Africa

- Mr Frikkie van der Merwe for guidance, assistance and mentorship
- Prof George van Schoor for guidance and mentorship
- Prof Kenny Uren for guidance and mentorship
- Mrs Louise Cilliers for financial and transportation assistance
- Mr Shawn Liebenberg for statistical analysis and assistance

Abstract

The inference modelling of a continuous industrial ammonium nitrate fluidised bed granulator was investigated. Fluidised bed granulators are difficult to model and control due to a large number of operating variables and long sample analysis time.

This study aims at developing inference models using the multiple linear regression (MLR) modelling technique to predict the output variables. These MLR models are qualitatively compared to the artificial neural network (ANN) modelling technique.

The granulation process starts with seed particles that are fluidised by air. These fluidised particles are sprayed with a solution of ammonium nitrate and water to induce granule growth. The final particles exit the granulator and are sieved to form the final product. Different literature sources obtained correlations between most of the operating variables concerning the output variables. Some authors obtained significant regression and ANN models that provided accurate predictions.

A screening phase is conducted to determine the significance of the chosen operating variables. The complete randomise experimental design is used where each operating variable is varied randomly from each run to produce unbiased data and lower the number of data points required for model development.

The input variables include the fluidising air flow rate, fluidising air temperature, spray liquid flow rate, spray liquid temperature, spray liquid concentration, seed particle size and the seed particle size distribution slope. The output variables consist of the production rate, recycle ratio, efficiency, product porosity, product circularity, product mean particle size, product particle size distribution slope, granulator mean particle size and granulator particle size distribution slope.

The correlations between the operating and output variables are determined using the Spearman's rho correlation technique. It is concluded that the spray liquid variables had the strongest correlations with the output variables and are therefore important variables for the MLR main effect models.

Accurate main effects (MLR-M) and interaction effects (MLR-I) multiple linear regression models were developed with a fair performance. The addition of the two-way interactions increased the accuracy and performance of the MLR-I models. The adjusted coefficient of multiple determination was used to evaluate the addition of these variables. The MLR-I models were compared to the ANN models that included all the independent variables (ANN-

I). The MLR-I models performed in some cases better than the ANN-I due to undertraining from the ANN-I models.

The production rate, recycle ratio, efficiency and granulator particle slope models were accurate enough for prediction purposes. Future work can include the optimisation of the ANN models and the investigation of additional variables for inference modelling, e.g. the bed height, bed density and granulator humidity. A control system using the developed models can also be developed and evaluated on the plant.

Keywords: Statistical modelling, multiple linear regression, artificial neural networks, fluidised bed granulation, inference measurement

Table of Contents

Acknowledgements ii

Abstract iii

List of Symbols viii

Abbreviations ix

List of figures x

List of tables xiii

Chapter 1. Introduction 1

1.1 Background to this study 1

1.2 Focus of this study 2

1.3 Objectives and scope of this work 3

Chapter 2. Literature review 4

2.1 Granulation process 5

2.1.1 Fluidised bed granulation 6

2.1.2 Granule growth mechanisms 7

2.1.3 Types of FBG 9

2.1.4 Influence of operating variables on the FBG operation 10

2.1.4.1 Fluidising air flow rate 10

2.1.4.2 Fluidising air temperature 11

2.1.4.3 Spray liquid flow rate 12

2.1.4.4 Spray liquid temperature 12

2.1.4.5 Spray liquid concentration 12

2.1.4.6 Seed particle size 13

2.1.4.7 Summary of influences 13

2.2 Modelling of the FBG 15

2.2.1 Previous studies 16

2.2.1.1 Multiple linear regression 17

2.2.1.2 Artificial neural networks 18

2.2.2 Regression models 19

2.2.3 Artificial neural networks 20
2.3 Critical literature review 24
Chapter 3. Experimental Procedures 25
3.1 Experimental setup 25
3.1.1 Materials used 25
3.1.2 Process flow diagram 26
3.1.3 Process sampling 27
3.1.4 Operating variables 27
3.1.5 Definition of output variables 29
3.1.5.1 Production rate 29
3.1.5.2 Recycle ratio 29
3.1.5.3 Efficiency 29
3.1.5.4 Particle size distribution and circularity 30
3.1.5.5 Porosity 32
3.2 Experimental procedure 35
3.2.1 Phase one 35
3.2.2 Phase two 36
3.3 Model development 37
3.3.1 Spearman's rho correlation 37
3.3.2 Multiple linear regression model development 38
3.3.3 Artificial Neural Network model development 39
3.3.4 Model performance evaluation 41
Chapter 4. Results and discussion 44
4.1 Results from phase one 44
4.1.1 Spearman's rho matrix 45
4.1.2 Multiple linear regression 51
4.2 Results from phase two 57
4.2.1 Spearman's rho matrix 57
4.2.2 Multiple linear regression 65

4.2.3 Comparison between MLR and ANN models 75
Chapter 5. Conclusions and recommendations 87
5.1 Conclusions 87
5.2 Recommendations 88
Chapter 6. References 90
Appendix A - Calculations 95
Appendix B - Experimental Planning 99
Appendix C - Artificial neural network comparison 102
Appendix D - Results 113
Appendix E - Data 120

List of Symbols

Symbol	Description
A	Projected area of the particle
b_i	Regression coefficient of independent variable i
d_{25}	First quartile particle size
d_{50}	Mean particle size
d_{75}	Third quartile particle size
$\boldsymbol{e}$	Error matrix between predicted and observed values
ϵ	Random error in Y
f	Activation function
i	Artificial neural network layer input neuron
$\boldsymbol{J}$	Jacobian matrix
j	Artificial neural network layer
k	Number of independent variables
m	Slope
N	Total number of weights
n	Sample size
o	Artificial neural network layer output neuron
P	Perimeter of the particle
Ψ	Circularity
R^2	Coefficient of multiple determination
ρ	Spearman's rho
S_{b_j}	Standard error of the regression coefficient b_j
S_{XY}	Standard error of estimate
s_i	Standard deviation of variable i
X	Independent variable
$\bar{X}$	Mean of independent variable
Y	Observed value of Y
$\bar{Y}$	Mean value of Y
$\hat{Y}$	Predicted value of Y
Y_i	Inference value of Y
Y_p	Primary (actual) value of Y
W	Weight

Abbreviations

Abbreviation	Description
ACD	Adjusted coefficient of multiple determination
AN	Ammonium nitrate
ANN	Artificial neural network
ANN-I	Artificial neural network All variables models
ANN-M	Artificial neural network Main effect variables models
BET	Brunauer Emmett Teller
CMD	Coefficient of multiple determination
CV	Coefficient of variance
CT	Computerised tomography
EFF	Efficiency
FAF	Fluidising air flow rate
FAT	Fluidising air temperature
FBG	Fluidised bed granulator
GPS	Granulator mean particle size
GSL	Granulator (linearized PSD) particle slope
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MLR	Multiple linear regression
MLR-I	Multiple linear regression Interaction models
MLR-M	Multiple linear regression Main effect models
MSE	Mean square error
PCR	Product circularity
PLS	Partial least squares
POR	Product porosity
PPS	Product mean particle size
PSD	Particle size distribution
PSL	Product (linearized PSD) particle slope
RER	Recycle ratio
RMSE	Root mean square error
RUR	Production rate
SEE	Standard error of estimate
SLC	Spray liquid concentration
SLF	Spray liquid flow rate
SLT	Spray liquid temperature
SPS	Seed mean particle size
SSL	Seed (linearized PSD) particle slope

List of figures

Figure 1-1: Block diagram for inference control adapted from Seborg et al. (2011, p.297) 2
Figure 1-2: Scope of work 3
Figure 2-1: Continuous fluidised bed granulator Qiu et al. (2009, p.704) 6
Figure 2-2: Granule growth mechanisms adapted from Sahoo (2012) 7
Figure 2-3: SEM of a layered and an agglomerated granule captured at the LEM of the North-West University 8
Figure 2-4: Cosine performance example 16
Figure 2-5: Cosine regression example 16
Figure 2-6: Common activation functions for artificial neural networks adapted from Ayat & Pour (2014) 22
Figure 2-7: The error-valley problem adapted from Kriesel (2007, p.53) 23
Figure 3-1: Process flow diagram with points of interest 26
Figure 3-2: Cumulative particle size distributions with different slopes adapted from Rhodes (2008, p.8) 30
Figure 3-3: Differential frequency distribution with different slopes adapted from Rhodes (2008, p.4) 31
Figure 3-4: Oil absorption setup 32
Figure 3-5: Porosity results from both techniques 34
Figure 3-6: Monotonic and non-monotonic functions adapted from Hauke & Kossowski (2011) 37
Figure 4-1: Phase 1 MLR main effects RUR model performance 51
Figure 4-2: Phase 1 MLR main effects RER model performance 52
Figure 4-3: Phase 1 MLR main effects EFF model performance 52
Figure 4-4: Phase 1 MLR main effects PPS model performance 53
Figure 4-5: Phase 1 MLR main effects GPS model performance 53
Figure 4-6: Phase 1 MLR main effects PCR model performance 54
Figure 4-7: Phase 1 MLR main effects PSL model performance 54
Figure 4-8: Phase 1 MLR main effects GSL model performance 55
Figure 4-9: Phase 1 MLR interaction effects POR model performance 56
Figure 4-10: Phase 2 MLR-M RUR model regression 66
Figure 4-11: Phase 2 MLR-I RUR model regression 66
Figure 4-12: Phase 2 MLR-M RER model performance 67
Figure 4-13: Phase 2 MLR-I RER model performance 68
Figure 4-14: Phase 2 MLR-M EFF model performance 69
Figure 4-15: Phase 2 MLR-I EFF model performance 69

Figure 4-16: Phase 2 MLR-M PPS model performance 70
Figure 4-17: Phase 2 MLR-I PPS model performance 70
Figure 4-18: Phase 2 MLR-I POR model performance 71
Figure 4-19: Phase 2 MLR-M PCR model performance 72
Figure 4-20: Phase 2 MLR-I PCR model performance 72
Figure 4-21: Phase 2 MLR-M PSL model performance 73
Figure 4-22: Phase 2 MLR-I PSL model performance 73
Figure 4-23: Phase 2 MLR-M GSL model performance 74
Figure 4-24: Phase 2 MLR-I GSL model performance 74
Figure 4-25: Phase 2 MLR-I RUR model performance 76
Figure 4-26: Phase 2 ANN-I RUR model performance 76
Figure 4-27: Phase 2 MLR-I RER model regression 77
Figure 4-28: Phase 2 ANN-I RER model regression 77
Figure 4-29: Phase 2 MLR-I EFF model performance 78
Figure 4-30: Phase 2 ANN-I EFF model performance 79
Figure 4-31: Phase 2 MLR-I PPS model performance 79
Figure 4-32: Phase 2 ANN-I PPS model performance 80
Figure 4-33: Phase 2 MLR-I GPS model regression 81
Figure 4-34: Phase 2 ANN-I GPS model regression 81
Figure 4-35: Phase 2 MLR-I POR model regression 82
Figure 4-36: Phase 2 ANN-I POR model regression 82
Figure 4-37: Phase 2 MLR-I PCR model performance 83
Figure 4-38: Phase 2 ANN-I PCR model performance 83
Figure 4-39: Phase 2 ANN-I PCR model regression 84
Figure 4-40: Phase 2 MLR-I PSL model performance 85
Figure 4-41: Phase 2 ANN-I PSL model performance 85
Figure 4-42: Phase 2 MLR-I GSL model performance 85
Figure 4-43: Phase 2 ANN-I GSL model performance 86
Figure A-1: t critical values for confidence intervals taken from Berenson et al. (2012, p.800) 96
Figure A-2: An illustration of an artificial neural network topography adapted from Wilamowski & Irwin (2011, p.10). 97
Figure C-1: Phase 2 ANN-M RUR model performance 103
Figure C-2: Phase 2 ANN-I RUR model performance 103
Figure C-3: Phase 2 ANN-M RER model performance 104
Figure C-4: Phase 2 ANN-I RER model performance 104
Figure C-5: Phase 2 ANN-M EFF model performance 105

Figure C-6: Phase 2 ANN-I EFF model performance 106
Figure C-7: Phase 2 ANN-M PPS model training error 107
Figure C-8: Phase 2 ANN-M PPS model validation error 107
Figure C-9: Phase 2 ANN-M PPS model performance 108
Figure C-10: Phase 2 ANN-I PPS model performance 108
Figure C-11: Phase 2 ANN-M PCR model regression 109
Figure C-12: Phase 2 ANN-I PCR model regression 109
Figure C-13: Phase 2 ANN-M PSL model regression 110
Figure C-14: Phase 2 ANN-I PSL model regression 110
Figure C-15: Phase 2 ANN-M GSL model performance 111
Figure C-16: Phase 2 ANN-I GSL model performance 111

List of tables

Table 2-1: Comparison of different FBGs adapted from Srivastava & Mishra (2010) 10
Table 2-2: Summary of the observations of the influences of operating variables of a FBG. 14
Table 3-1: Materials used 25
Table 3-2: Summary of the process variables 27
Table 3-3: Statistical results from porosity methods 33
Table 3-4: Repeatability results for phase 2 36
Table 3-5: List of interaction combinations 39
Table 3-6: Neural network training parameter values Beale et al. (2014, p.327) 40
Table 3-7: Difference between MAE and RMSE adapted from Willmott & Matsuura (2005) 43
Table 4-1: Frist Spearman's rho matrix for phase 1 45
Table 4-2: Second Spearman's rho matrix for phase 1 46
Table 4-3: Third Spearman's rho matrix for phase 1 49
Table 4-4: Main effects multiple linear regression model results for phase 1 51
Table 4-5: Interaction effects multiple linear regression model results for phase 1 55
Table 4-6: Frist Spearman's rho matrix for phase 2 58
Table 4-7: Second Spearman's rho matrix for phase 2 59
Table 4-8: Third Spearman's rho matrix for phase 2 62
Table 4-9: Multiple linear regression results for phase 2 65
Table 4-10: Phase 2 MLR model comparison with the ANN models 75
Table B-1: Abbreviations of process variables 99
Table B-2: Phase one variable ranges 99
Table B-3: Phase one experimental plan 99
Table B-4: Phase two variable ranges 100
Table B-5: Phase two spray liquid temperatures for different concentrations 100
Table B-6: Phase two experimental plan 101
Table C-1: Artificial neural network results for phase 2 102
Table D-1: Spearman's rho matrix for phase 1 113
Table D-2: Spearman's rho p-values for phase 1 114
Table D-3: Spearman's rho matrix for phase 2 115
Table D-4: Spearman's rho p-values for phase 2 116
Table D-5: MLR main effects model coefficients for phase 1 117
Table D-6: MLR interaction effects model coefficients for phase 1 118
Table D-7: MLR main effects model coefficients for phase 2 118
Table D-8: MLR interaction effects model coefficients for phase 2 119
Table E-1: Phase 1 dataset 120

Table E-2: Phase 2 dataset..... 122

Chapter 1. Introduction

1.1 Background to this study

Fluidised bed granulation is a wet granulation technique that is widely used in the agriculture, pharmaceutical, food and chemical industry. Granulation is one of the main processes in the manufacturing of fertiliser. This technique is used to produce granules by fluidising seed particles and spraying a liquid product onto the surfaces of the fluidised particles (Biswal 2011). Granulation is utilised for particle growth and for improvement of the material properties that includes the dissolution rate, bulk density and granule strength (Burggraeve et al. 2013; Srinivasakannan & Balasubramaniam 2003).

A fluidised bed granulator (FBG) has the advantage of various processes taking place in a single operating unit with the added advantages of low labour costs and relatively short process times. The mass and heat transfer is more contained with a uniform distribution due to operation in a single unit (Burggraeve et al. 2013; Srinivasakannan & Balasubramaniam 2003).

The granulation process involves three steps that consist of the wetting, growth and attrition of particles. During the wetting step, a liquid is atomised and sprayed into the fluidised bedchamber by means of a nozzle with pressurised air (Becher & Schlünder 1998). Spraying can be done from the top, side or bottom of the chamber (Wong et al. 2013).

Particle growth can follow two mechanisms depending on the operating conditions (Biswal 2011). The first is layering where the fine droplets form a layer on the surface and dries before collisions occur. The second mechanism is agglomeration where wetted particles collide with other particles to form a granule. A liquid bridge forms and hardens creating an agglomerate (Becher & Schlünder 1998). Particles also collide with the inner wall of the fluidised bed chamber resulting in attrition to occur, forming smaller, lumpier particles. The last step occurs independent of the growth mechanism (Sahoo 2012).

1.2 Focus of this study

FBGs are difficult to model due to the complexity of the different processes occurring in the granulator (Cameron et al. 2005). Different process variables have an influence on the operation and product quality that makes control difficult. Different literature sources use statistical correlation methods to determine the relationships between the variables (Sahoo 2012; Wong et al. 2013).

Due to the many difficulties of wet granulation, granulation itself has been considered more of an art than a science. Explaining and predicting the behaviour of granulation is one of many problems along with large recycle ratios and lack of quality control (Iveson et al. 2001).

The quality of the granules is determined from laboratory analysis that is time consuming. Burggraeve et al. (2013) suggest that the granulation process can be better controlled using in-line quality devices or inference sensors. A few of the FBG variables are measured during the granulation process and can be used with multidimensional modelling techniques to predict the output variables of the FBG. Focussing on developing models to predict the output variables in real-time, using the operating variables in an inference model, will increase the controllability of the FBG process (Zhu et al. 2011). A block diagram for inference control is given in Figure 1-1 where an inference model uses process variables to predict the quality variable, Y, in real-time. The predicted value, Y_i, is used for control purposes. The inferential model is adapted from the actual quality variable, Y_p, obtained from the laboratory analysis.

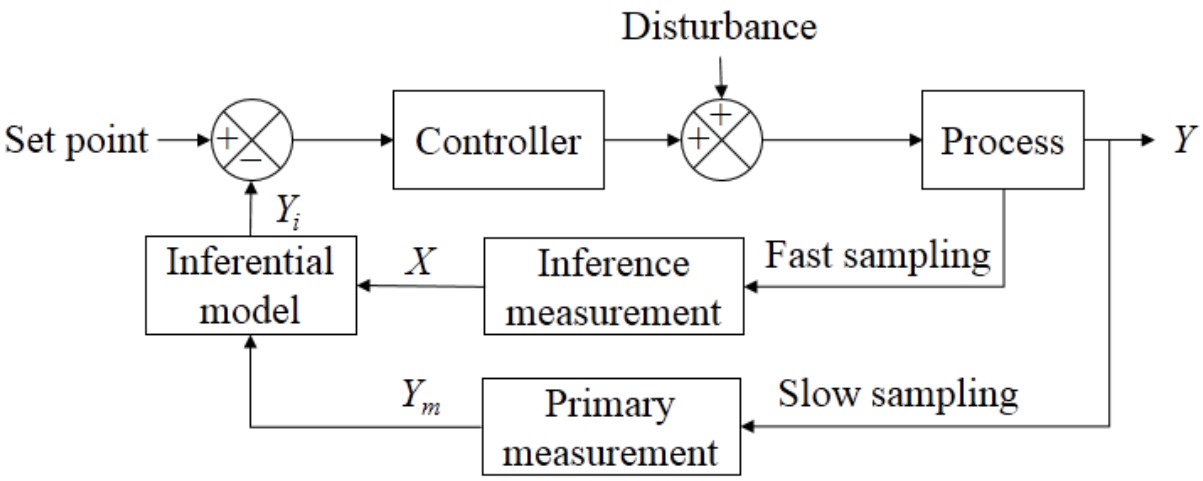


Figure 1-1: Block diagram for inference control adapted from Seborg et al. (2011, p.297)

1.3 Objectives and scope of this work

The primary objective of this study is to obtain statistical significant relationships between the operating variables and measurable output variables of a FBG. Statistical models will be developed, assessed and validated for inference measurement purposes. A secondary objective is to compare the performance and accuracy of the statistical models with elementary artificial neural network models. The scope for this study is shown in Figure 1-2.

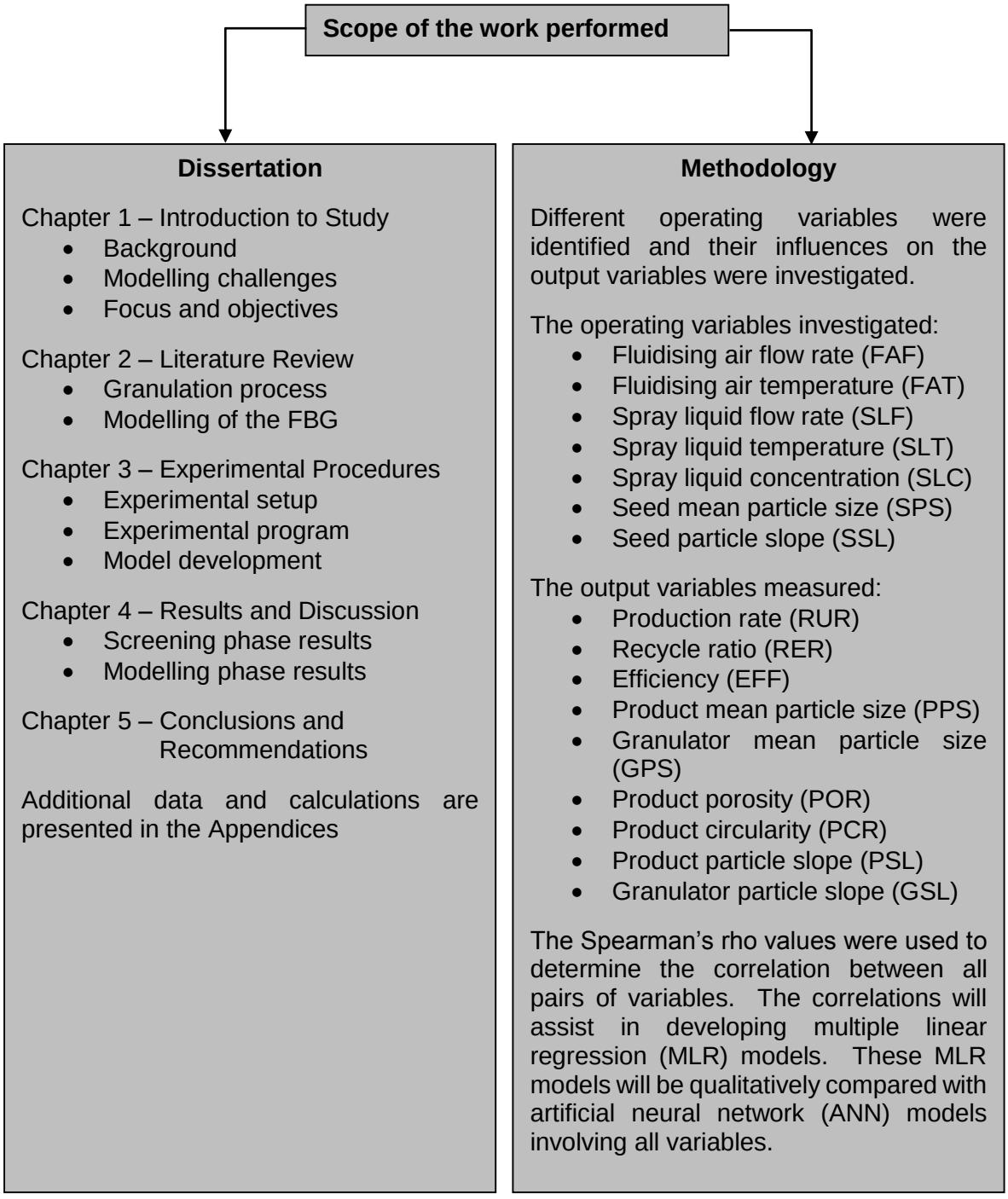


Figure 1-2: Scope of work

Chapter 2. Literature review

Fluidised bed granulation is one of many granulation processes that are not fully understood and is operated inefficiently by means of popular practice. The granulation process presents some challenges especially in the area of multiscale modelling, which consist of modelling of the particles and process, with the aim of optimisation and control (Cameron et al. 2005).

A thorough understanding of the granulation process and influence of the operating parameters on the behaviour of a fluidised bed granulator (FBG) is necessary for model development (Ziyani & Fatah 2014).

Granulation can be divided into wetting of seed particles, particle growth and lastly drying of final product. FBGs are single operating units in which all these processes occur, making it ideal for granulation. Fluidised bed granulation is preferred over conventional techniques due to its low labour and operational costs (Iveson et al. 2001; Qiu et al. 2009, p.701).

The control and operation of a FBG is difficult due to the integrated process with a large number of operating variables and disturbances causing unstable operation such as defluidisation (Cameron et al. 2005). Defluidisation can occur due to insufficient air flow or the formation of too large granules, resulting in the collapsing of the granulator bed and halting the production of granules (Srinivasakannan & Balasubramaniam 2003). In studies conducted by Becher & Schlünder (1998), Biswal (2011), Fries et al. (2014), Sahoo (2012) and Srinivasakannan & Balasubramaniam (2003), the influences and relationships between different operating and output variables were investigated. Different modelling approaches were considered with the most common approach being the black-box approach. Different useable empirical models were developed using statistical approaches (Burggraeve et al. 2013; Murtoniemi et al. 1994; Patel et al. 2010).

The granulation process, the FBG process description and different FBG types will firstly be discussed in this chapter. The observations and conclusions from different authors on the influences of the different operating variables on the output variables will be discussed in section 2.1.4. The modelling of a fluidised bed granulator will be discussed in section 2.2. The section will consist of the different modelling approaches from previous studies, followed by a detail description on these modelling techniques.

2.1 Granulation process

There is a lack of formal methodology regarding the design and operation of a granulator. This results in industrial plants experiencing surging, cyclic behaviour and unpredictable product quality. Complex disturbances, long lag times and large recycle ratios, some in the order of 5:1, lead to unstable operations and deviations from the desired outputs (Iveson et al. 2001; Sahoo 2012; Ziyani & Fatah 2014).

Granulation is the process wherein granules are bound together to form a single, larger granule from which the original granules can still be identified. There exist two types of granulation referred to as dry and wet granulation. Dry granulation is the formation of granules using mechanical compaction. The powder is compressed into a pellet usually in a process called slugging. This method is used when the powder is sensitive to moisture and heat (Biswal 2011; Moraga et al. 2015; Sahoo 2012).

A spray liquid is used to induce granule growth during wet granulation. The spray liquid consists out of a binder and solution or melt of the product material. The binder is used to bind the granules together using viscous and capillary forces until a permanent bond is formed. A drying step is required for wet granulation, which involves drying equipment. The binder material should be volatile in order to be evaporated during the drying phase. Water is commonly used as a binder from ecological and economical viewpoints although it may take longer to evaporate compared to other organic solvents (Iveson et al. 2001; Ziyani & Fatah 2014).

The granulation process is used to improve the final properties of the product that includes the circularity, bulk density and porosity. Circularity is a measurement of how close the projected area of a granule is to a circle. This property is associated with the flow ability and handling of granules. Granulation decreases the dustiness of the initial product, resulting in reducing product losses and also explosion and inhalation risks. The dissolution rates are improved by producing a more porous product. The product would also have a high proportion of the original granules' surface area, which makes the product ideal for catalytic applications (Iveson et al. 2001; Pashminehazar et al. 2016).

2.1.1 Fluidised bed granulation

A FBG is a wet granulation technique that involves spraying air suspended seed particles with a liquid. The liquid consists of a solution or melt of the product material along with a binder that is sprayed via a nozzle as shown in Figure 2-1. The particles are wetted and collide with one another forming agglomerates resulting in particle growth (Sahoo 2012).

A FBG is referred to as a one-pot process where blending, wet granulation and drying all occur. Fluidisation induces mixing resulting in improved mass and heat transfer between the air, liquid and solids to such an extent that near zero concentration and thermal gradients are achieved. This gives FBGs the advantage over conventional methods by saving time, labour and operational costs (Burggraeve et al. 2013; Qiu et al. 2009, p.701,704; Ziyani & Fatah 2014).

Quenching, a defluidisation phenomenon, is one main disadvantage of FBG as granulation technique. Quenching can be divided into wet and dry quenching. Wet quenching occurs when large wet lumps form due to inadequate drying. Operation with excessive liquid, i.e. high liquid flow rates, causes overwetting of granules. The granules stick together forming large wet lumps causing the bed to defluidise (Smith & Nienow 1983; Wong et al. 2013).

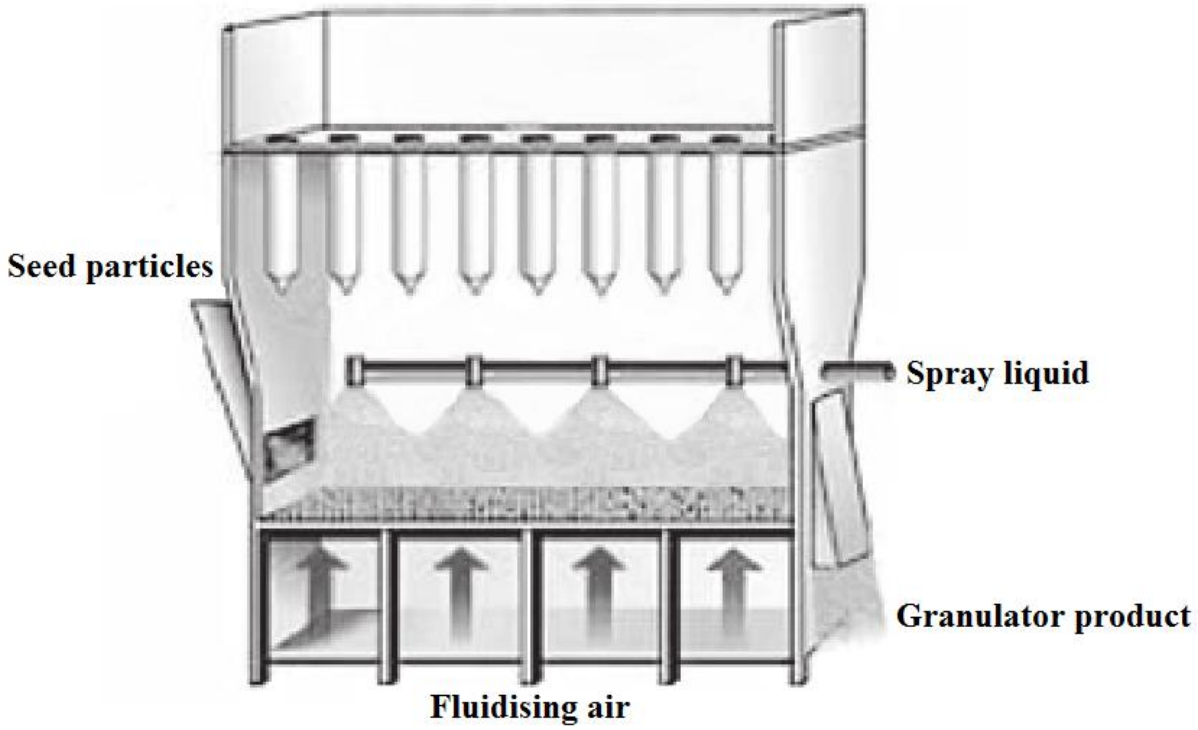


Figure 2-1: Continuous fluidised bed granulator Qiu et al. (2009, p.704)

Agglomeration may become uncontrollable at certain conditions, resulting in the formation of large agglomerates. Granulation will occur up to a point where the fluidising air will become insufficient and the bed will start to defluidise. This phenomenon is referred to as dry quenching (Becher & Schlünder 1998; Smith & Nienow 1983).

Controlling a FBG proves to be difficult due to a lack of understanding of the process and its behaviour (Cameron et al. 2005). Fluidised bed granulation is proven to be a complex, multidimensional process with different variables that influences the granulation process and stability of the process (Srinivasakannan & Balasubramaniam 2003). The process suffers from unstable operation due to some disturbances. Some variables are difficult to measure online, resulting in difficulty to determine the effect of such disturbances in real-time (Burggraeve et al. 2013; Patel et al. 2010). Deviations in output variables are amplified due to long lag times which also cause unstable operation leading to defluidisation (Zhang et al. 2000; Ziyani & Fatah 2014).

2.1.2 Granule growth mechanisms

The granulation growth process is divided into wetting and nucleation, granule growth and lastly attrition and breakage as shown in Figure 2-2 (Iveson et al. 2001).

The process starts with solid feed material, referred to as seed particles, entering the granulator, which are fluidised by hot air. Wetting and nucleation is the first process to occur where dry particles collide with liquid droplets forming wetted particles. The rates at which the particles are wetted are determined by different nozzle positions, droplet sizes and binder addition rates (Sahoo 2012).

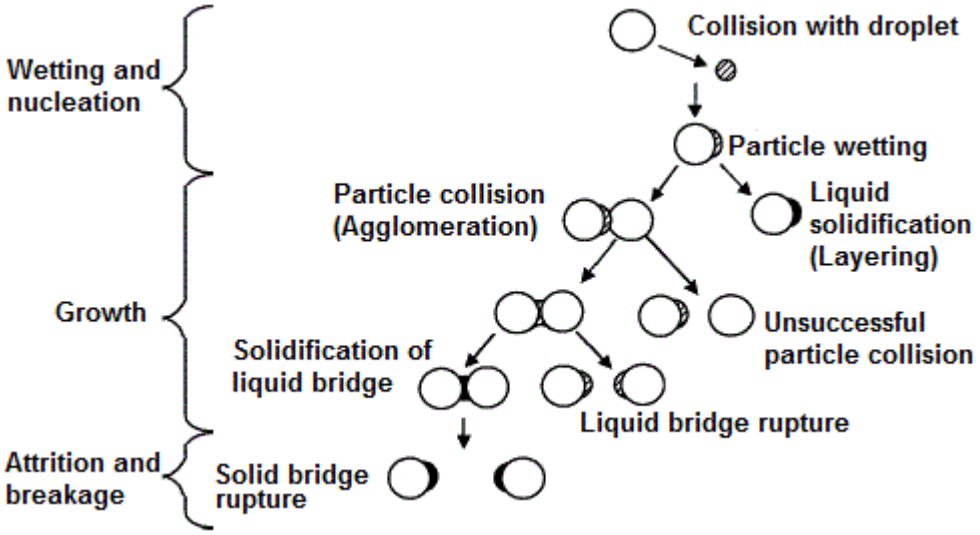


Figure 2-2: Granule growth mechanisms adapted from Sahoo (2012)

Growth can follow one of two mechanisms called layering and agglomeration. The growth mechanism depends on the operating conditions of the FBG and the physicochemical properties of the material (Biswal 2011; Srinivasakannan & Balasubramaniam 2003). A Quanta FEG 250 Environmental Scanning electron microscope (SEM) was used to capture the surface morphology of different ammonium nitrate granules produced by each mechanism as shown in Figure 2-3. A 126 acceleration voltage of 30kV was used while improving the image by lightly covering the granules with a gold 128 layer at the Laboratory of Electron Microscopy (LEM) of the North-West University.

Layering is one route that a wetted particle can follow. The spray liquid solidifies forming a layer around the particle resulting in a uniform spherical granule. The rate of solidification is in the order of milliseconds while the collision between the particles and droplets are in the order of microseconds. The large difference in orders between the two rates results in a low probability of layering (Biswal 2011; Sahoo 2012; Srinivasakannan & Balasubramaniam 2003).

The most probable route will be the collision between wetted particles forming agglomerates. Agglomerates can also be formed from collisions between wet particles and dry particles. Liquid bridges hold the newly formed agglomerates together. If the binder is not sufficient to hold the particles together, unsuccessful agglomeration will occur and the particles will separate (Iveson et al. 2001; Sahoo 2012).

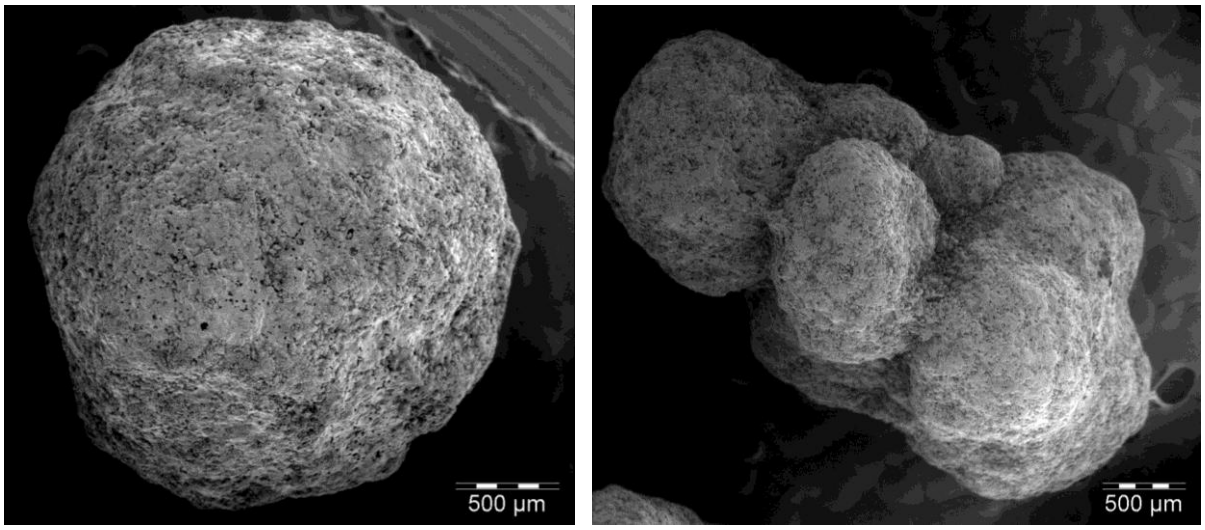


Figure 2-3: SEM of a layered and an agglomerated granule captured at the LEM of the North-West University

Drying of the agglomerates starts to occur and the liquid bridges starts to solidify. The nozzle flow rate, bed height and the fluidising temperature influences the drying zone in a FBG (Becher & Schlünder 1998). Attrition and breakage occurs where the agglomerates break due to collisions between dry granules or collisions with the granulator wall. Some agglomerates break due to formation of weak solid bridges, resulting in bridge rapture. Some agglomerates withstand attrition and form granules. The number of granules formed depends on the solidification rate and binder used (Qiu et al. 2009, p.706; Sahoo 2012).

2.1.3 Types of FBG

FBGs can be classified in three different types based on their spraying location. Each type favours different product properties due to the different positions of the spraying nozzles. Top spraying is one of the most common types used in the granulation industry. Top spraying FBGs favour the agglomeration mechanism and produces a porous granule that has a good dispersibility, but is friable (Qiu et al. 2009, p.702; Srivastava & Mishra 2010).

Bottom spraying FBGs favour the layering mechanism with a higher coating efficiency compared to top spraying FBGs. A bottom spraying FBG cannot be operated at high liquid flow rates without the risk of overwetting the granules, resulting in wet quenching and defluidisation to occur. They are therefore operated at lower liquid flow rates producing denser, smoother and uniform granules at a lower production rates (Qiu et al. 2009, p.702; Wong et al. 2013).

The last type is tangential spraying that has a combination of the properties from both the top and bottom spraying types. Tangential spraying systems can be operated at higher liquid flow rates. This reduces the process time and increases the production rate. The granules produced are less porous compared to top spraying granules but more spherical and rigid (Srinivasakannan & Balasubramaniam 2003; Wong et al. 2013). Srivastava & Mishra (2010) compared all three types of FBGs and the conclusions are summarised in Table 2-1. The FBG investigated in this project was a tangential spraying FBG, making the right column of the table important. All the product considerations are highly desirable along with accessibility of the process equipment, i.e. the nozzles, making the tangential spraying FBG ideal for continuous production.

Table 2-1: Comparison of different FBGs adapted from Srivastava & Mishra (2010)

	Top spray	Bottom spray	Tangential spray
Process consideration			
Simplicity	Highly desirable	Desirable	Least desirable
Nozzle access	Highly desirable	Least desirable	Desirable
Scale-up issues	Highly desirable	Desirable	Least desirable
Product considerations			
Surface morphology	Least desirable	Highly desirable	Highly desirable
Coating uniformity	Desirable	Highly desirable	Highly desirable
Layering efficiency	Least desirable	Highly desirable	Highly desirable
Product coating capacity	Desirable	Highly desirable	Highly desirable

2.1.4 Influence of operating variables on the FBG operation

Fluidised bed granulation is a complex process with many influential process variables such as the fluidising conditions, atomising spray conditions and the physicochemical properties of the spray liquid. There exists a significant relationship between the operating variables and the product properties due to the highly interconnected process. This makes controlling of a FBG difficult (Patel et al. 2010; Srinivasakannan & Balasubramaniam 2003).

2.1.4.1 Fluidising air flow rate

The fluidising air flow rate (FAF) influences both the heat and mass transfer kinetics of the process and the dynamics of the particles. These include the collision and mixing rate of the particles and the bed height. It is important to operate the FBG at a certain fluidised air flow rate to keep the particles fluidised otherwise the bed may collapse. (Fries et al. 2014).

Biswal (2011), Rambali et al. (2001) and Sahoo (2012) have found that an increase in the FAF increases the attrition rate of the particles. This resulted in the reduction in the agglomeration rate and the formation of smaller granules and dust.

Smith & Nienow (1983) found that the layer growth mechanism becomes dominant at higher fluidisation velocities. Increasing the fluidising air causes the bed particles to mix more vigorously resulting in more collisions where the agglomerate bridges break more rapidly. The authors concluded that smaller granules are formed at higher fluidising velocities.

Srinivasakannan & Balasubramaniam (2003) observed that operating at higher FAF allow FBGs to operate at higher liquid flow rates without the risk of wet quenching to occur. Fries et al. (2014) concluded that an increase in the flow rate of the fluidising air will result in a higher

drying potential which results in a shorter drying period. Fewer wetted particles will be available for collision that reduces the agglomeration rate. Operating at a higher fluidising flow rate increases the agitation resulting in an increase in the attrition and breakage of granules. The authors concluded that an increase in the flow rate would result in a decrease in the product particles size.

2.1.4.2 Fluidising air temperature

Fluidised air enters the granulator at a defined temperature where it is used to heat the seed particles. This stored heat in the seed particles is used in the beginning of the drying phase to evaporate most of the solvent in the spray liquid (Becher & Schlünder 1998).

Fluidising air temperature (FAT) is important for the operation of the FBG. Operating at a low temperature may cause lumps to form, resulting in a collapsed fluidised bed. The FAT is a key variable in order to avoid wet and dry quenching (Ziyani & Fatah 2014).

Becher & Schlünder (1998), Biswal (2011), Sahoo (2012) and Ziyani & Fatah (2014) noticed that increasing the air temperature results in an increase in the evaporation rate and the formation of smaller granules. Srinivasakannan & Balasubramaniam (2003) concluded that the FAT has no influence on the production rate. The authors also observed that a FBG could be operated at higher liquid flow rates without the risk of wet quenching when higher FATs are used.

Singh et al. (2011) found that higher air temperatures will result in higher evaporation rates and a shorter drying phase. The sudden evaporation will result in an increase in friable granules. The solid bridges will break more easily during collisions, resulting in a decrease in granule size.

Fries et al. (2014) concluded that increasing the process temperature, a consequence of increasing the FAT, will result in a faster evaporation of the water in the spray solution. A lower humidity will result in fewer wetted particles available for collision, lowering the agglomeration rate. Increasing the temperature resulted in the formation of unstable bridges due to the increased drying rate, decreasing the overall agglomeration.

2.1.4.3 Spray liquid flow rate

The spray liquid flow rate (SLF) has a large influence on the throughput and particle growth rate of a FBG. Operating at a higher flow rate requires a higher agitation and mixing rate, otherwise overwetting can occur, causing the bed to defluidise (Fries et al. 2014).

Becher & Schlünder (1998), Sahoo (2012) and Srinivasakannan & Balasubramaniam (2003) found that an increase in the SLF results in the formation of larger liquid droplets. This increases the collision rate of particles with liquid droplets resulting in an increase in wetted particles. A larger number of wetter particles result in the production of larger granules, increasing the agglomeration rate.

Fries et al. (2014) and Wong et al. (2013) concluded that increasing the spray liquid flow rate will result in an increase in the relative humidity of the granulator. This is favourable for agglomeration, but poses an increased risk of defluidisation. All authors also observed that an increase in the SLF results in an increase in liquid droplet size, leading to an increase in granule growth rate.

2.1.4.4 Spray liquid temperature

Sahoo (2012) found that an increase in the temperature of the spray liquid results in a decrease in the viscosity. A decrease in viscosity leads to a decrease in the formation of agglomerates due to the formation of weak liquid bridges. Less energy is required to break the liquid droplets if the viscosity is low, leading to smaller droplets forming smaller granules.

2.1.4.5 Spray liquid concentration

Operating at a higher spray liquid concentration (SLC) results in an increase in the amount of solute sprayed and a decrease in the amount of solvent, i.e. water, which needs to be evaporated. Srinivasakannan & Balasubramaniam (2003) observed that at higher concentrations the growth rate of the granules increased. The authors concluded that higher solute concentrations favour the agglomeration growth mechanism. Sahoo (2012) noticed that higher concentrations result in a higher viscous liquid bonding, resulting in the formation of more stable agglomerates and an increase in the agglomeration rate.

Burggraeve et al. (2011) studied the influence of the SLC on the product size distribution. The authors found that the SLC variable had a positive effect on the product mean particle size,

d_{50}. An increase in the concentration resulted in larger particles being formed and a reduction in fine particles.

2.1.4.6 Seed particle size

A smaller seed mean particle size (SPS) increases the total surface area of the bed mass. Small particles can also lead to problems during fluidisation (Sahoo 2012).

Biswal (2011) and Srinivasakannan & Balasubramaniam (2003) found that operating at a larger mean SPS favours the layering growth mechanism. The growth rate of the layering mechanism is lower in comparison to the agglomeration mechanism. Srinivasakannan & Balasubramaniam (2003) concluded that the SPS has a major influence on the growth mechanism of the granules. Smaller seed particles are used if a higher throughput ratio is required. Smaller diameters favour agglomeration resulting in a higher growth rate.

2.1.4.7 Summary of influences

Table 2-2 summarises the influences of the different operating variables obtained from literature. Various authors have investigated most of the operating variables while the SLT variable did not receive much attention. Most of the FBGs investigated were laboratory scale granulators operated in batch. Not much work has been done on continuous FBGs especially on industrial scale granulators.

Table 2-2: Summary of the observations of the influences of operating variables of a FBG.

Variable	Source	Observations
Fluidising air flow rate	Biswal (2011), Rambali et al. (2001), Sahoo (2012) and Smith & Nienow (1983)	Increasing the FAF increases attrition rate. Reduction of agglomerates and formation of smaller granules.
	Fries et al. (2014)	Increasing the FAF increases drying rate and attrition. Decreases particle diameter.
Fluidising air temperature	Becher & Schlünder (1998), Biswal (2011), Sahoo (2012) and Ziyani & Fatah (2014)	Increasing FAT increases evaporation rate. Formation of smaller granules.
	Fries et al. (2014)	Increasing FAT increase evaporation rate. Unstable bridges, decrease in agglomeration.
	Singh et al. (2011)	Increasing FAT increase evaporation rate. Increase in number of friable granules. Decrease in granule size
	Srinivasakannan & Balasubramaniam (2003)	Increasing FAT increases the liquid flow rate capacity. FAT has no effect on the growth rate.
Spray liquid flow rate	Becher & Schlünder (1998), Sahoo (2012) and Srinivasakannan & Balasubramaniam (2003)	Increase in SLF increases droplet size and collision rate. Larger granules and increase in agglomeration rate
	Fries et al. (2014) and Wong et al. (2013)	Increase in SLF increases relative humidity of granulator. Favours agglomeration, but high risk of defluidisation. Increase in droplet size and granule growth.
Spray liquid temperature	Sahoo (2012)	Increase in SLT decreases viscosity and decrease agglomeration.
Spray liquid concentration	Sahoo (2012)	Higher SLC higher viscous liquid bond. Stable agglomeration, increase in agglomeration rate.
	Srinivasakannan & Balasubramaniam (2003)	Higher SLC increase in growth rate. Agglomeration is favoured.
Seed particle size	Biswal (2011) and Srinivasakannan & Balasubramaniam (2003)	Decrease in SPS favours agglomeration. Growth rate is higher, desirable if higher throughput ratio is required.
	(Sahoo 2012)	Decrease in SPS increase in total surface area of bed mass. Smaller particles can cause fluidisation problems.

2.2 Modelling of the FBG

Fluidised bed granulation is a complex multidimensional process in which mixing, granulation and drying occur. The control of a FBG is difficult due to disturbances and influences on the process that are difficult to measure. Developing models as inference sensors can help in controlling and optimising a FBG (Burggraeve et al. 2013). A thorough understanding of the complexity of the operating variables and the influences of these variables on a FBG is required to achieve a consistent granulation process. Obtaining models describing the influences of the operating variables on the granulation process, as well as their interactions with each other, can improve the understanding, modelling and control of FBGs (Aleksic et al. 2014; Patel et al. 2010; Srinivasakannan & Balasubramaniam 2003).

Challenges are present in the multiscale modelling and control of a granulation process. Multiscale modelling consists of modelling the micro-scale that involves changes on the particle level and the macro-scale representing the process operation. Modelling approaches can be divided into white-box, grey-box and black-box approaches (Burggraeve et al. 2013; Cameron et al. 2005).

Modelling using fundamental physics and chemistry is considered a white-box approach. White-box models incorporate conservation aspects, such as thermodynamic principles, mass and heat transfer as well as particle growth. These models are more complex and time-consuming but are more flexible especially for scale-up and designing (Burggraeve et al. 2013; Cameron et al. 2005).

The grey-box approach includes the underlying understanding of the chemistry and physics involved in the process along with process data to develop a model. These models are also complex and time-consuming but provide semi-flexible models. Data fitting is still required for these models to be accurate (Burggraeve et al. 2013; Cameron et al. 2005).

Black-box approaches use empirical models or arbitrary functions to fit input-output data. These models can be used to get the optimal operating variables and conditions of the process that can be used for control purposes. The models can be developed quickly but are constricted to the design range. These models cannot be used for extrapolation beyond the known envelope and the results are hard to interpret. Black-box models are often considered, especially with complex processes such as granulation. Statistical models are some of the most commonly used black-box models which includes linear regression, partial least squares and artificial neural networks (Burggraeve et al. 2013; Cameron et al. 2005; Hill & Lewicki 2006, p.483; Zhu et al. 2011).

2.2.1 Previous studies

Various authors used different modelling techniques to determine the influences of the different operating variables on a FBG. Most authors used multiple linear regression (MLR) or artificial neural network (ANN) modelling techniques. A few authors, Cameron et al. (2005), Heinrich et al. (2003) and Palis et al. (2012), investigating the population balance technique and only Burggraeve et al. (2011) and Burggraeve et al. (2012) investigated the partial least squares (PLS) technique.

Most authors used the coefficient of multiple determination (CMD) as a performance criterion without investigating the error of the predicted results. The CMD only indicates the linear relationship between predicted and actual values. Predicted values that fit the actual values in a linear relation but with an offset, will still obtain a high CMD value. A good example is a cosine function representing the actual data and the same cosine function with a constant offset representing the model's prediction as shown in Figure 2-4. Plotting the actual versus predicted graph will indicate a CMD of one as shown in Figure 2-5; however, the data points do not fit each other. Reporting the error values give more significance to the performance evaluation of the models. The different performance evaluation methods are elaborated in section 3.3.4.

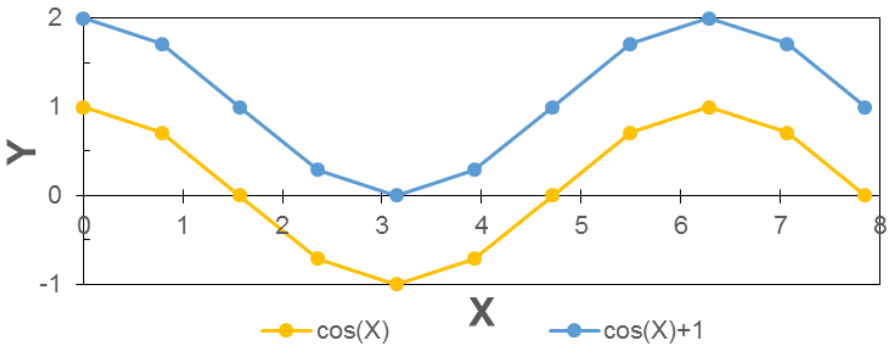


Figure 2-4: Cosine performance example

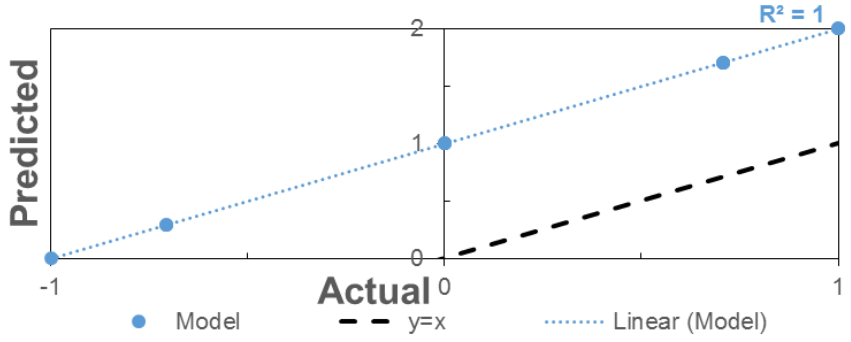


Figure 2-5: Cosine regression example

2.2.1.1 Multiple linear regression

Wong et al. (2013) investigated the influence of the binder addition, spray rate and distance between a top spraying nozzle and granulator bed on the product quality of a FBG, using quadratic MLR models. The authors could not find any statistical significant models for the process yield, number of fines produced and the circularity of the granules. Models were developed for the granule size, percentage lumps, size distribution span of the granules and the ratio of the tapped to bulk density. The span is the ratio of the difference between the ninetieth percentile particle size (d_{90}) and the tenth percentile particle size (d_{10}) with the mean particle size (d_{50}) as calculated by Equation 2-1.

$$\text{Span} = (d_{90} - d_{10})/d_{50} \qquad \textbf{Equation 2-1}$$

The linear and quadratic effects were found to be significant in the granule size model. It showed a fair prediction with a CMD of 88%. The binder addition was found as the most influential variable with a positive non-linear effect. The models for the percentage lumps and span of the granules showed regression performances of 76% and 86% respectively. The tapped to bulk density ratio had a weak predicted CMD of 54%. Not all the product output variables could be modelled, but the authors were still able to draw conclusions regarding the influences of the different independent variables on the product output variables and also whether interaction occurred between independent variables.

Ziyani & Fatah (2014) studied the influences of the FAF, FAT, SLF and spray liquid pressure on the granule properties to obtain the optimal operating conditions for a FBG. The authors developed MLR models with two-way interactions and found that the FAT had a significant effect on the particle size. The authors concluded that the spraying pressure should also be high to prevent formation of large droplets and wet quenching.

Rambali et al. (2001) investigated the use of MLR to model and optimise a FBG. The influences of the FAF, FAT, fluidising air humidity and SLF on the granule size were modelled using the quadratic polynomial MLR technique. Only statistical significant variables were included in the final model. The authors found the residuals of the MLR model to be normally distributed, concluding that the model fits adequately. The authors concluded that optimisation using the MLR models was successful with accurate predictions within the confidence intervals.

2.2.1.2 Artificial neural networks

Murtoniemi et al. (1994) compared ANN with MLR models using a FBG and investigated the influences of FAT, atomising air pressure and the binder addition rate on the mean particle size, d_{50}, and friability of the granules. The authors used optimised ANN models trained with the back-propagation training algorithm. The authors found the average percentage errors for their ANN models to be 15% and 8% for the d_{50} and friability models respectively. The average percentage errors for the MLR models were 17% and 11% respectively. The authors concluded that the ANN models performed more accurately and stated that acquiring more data along with obtaining the optimal topography and training parameters will lead to a more accurate model but could take a long time.

Aleksic et al. (2014) evaluated ANN models using the back-propagation training algorithm for modelling and predicting capabilities for a pharmaceutical FBG using Lactose and Gelucire as seed material and binder respectively. The binder and granulation time were used as the independent variables with the particle diameter, size distribution span, circularity and other product variables as the dependent variables. High CMD values were obtained ranging between 80% and 98%. The authors concluded that the ANN models proved successful as modelling technique for FBGs.

In another study Aleksic et al. (2015) studied the effect that the binder and spray air pressure has on the granule properties of a FBG. The authors compared PLS models with ANN models that were trained using the back-propagation training algorithm. Lactose was used as seed particles and sprayed with Gelucire as binder. Higher CMD of 99% were obtained. The authors observed that the ANN models had lower percentage prediction errors compared to the PLS models and concluded that both modelling techniques have a good predictability and accuracy.

Korteby et al. (2016) studied the prediction of the temperature profile inside a granulator using an ANN. The liquid flow rate and fluidising air temperature were varied and used as inputs. A CMD of 99% was found indicating the ANN's capability to predict the temperature profile of a FBG.

All the authors concluded that ANN's can be successfully used to model certain aspects of a FBG. There was; however, little investigation done in comparing inference models developed by both MLR and ANN techniques. Most of the authors only reported the CMD value without assessing the error performance as stated in section 2.2.1. A full comparison with detail evaluation is required to distinguish between the performances from each technique.

2.2.2 Regression models

There are different types of regression modelling techniques with the most common being the linear regression model which only considers one independent variable. Modelling multiple independent variables can be achieved using MLR (Berenson & Levine 1989, p.628).

MLR assesses the strength of the relationship between two or more independent variables and a dependent variable (Azadi & Karimi-Jashni 2016; Landau & Everitt 2004, p.92). MLR is popular due to its simplicity, easy implementation and interpretation (Vivaracho-Pascual et al. 2016). The general equation is given by Equation 2-2 where the predicted output, dependent variable, ($\hat{Y}$) is calculated as the product sum between all the independent variables (X_i) and their respected regression coefficients (b_i) with b_0 representing a constant value. The regression coefficients give an indication of the relationship between the independent and dependent variables. A positive value indicates that a positive relationship exist, while a negative value indicates a negative relationship (Berenson et al. 2012, p.580; Landau & Everitt 2004, p.109).

$$\hat{Y} = b_0 + \sum_{i=1}^{k} b_i X_i \qquad \textbf{Equation 2-2}$$

The regression coefficients are calculated by using the least squares method, which minimizes the sum of the squared errors between the model and the actual data points (Field 2009, p.209; Landau & Everitt 2004, p.93). MLR assumes that the relationship between a dependent and an independent variable is linear. This assumption does not restrict the technique and some valid models can be obtained for non-linear processes, if a sufficiently linear region exist or if it can be linearized (Hill & Lewicki 2006, p.395). MLR models are developed under the assumption that all the independent variables are independent from one another and that the dependent variable is normally distributed (Azadi & Karimi-Jashni 2016; Vivaracho-Pascual et al. 2016). These independences will be evaluated during model development.

The regression coefficients should be statistically significant to ensure accurate, usable MLR models. This is determined using a standard t-test with a 95% certainty. Equation 2-3 shows the calculation for a t-test where S_{b_i} represents the standard error of the regression coefficient b_i (Berenson et al. 2012, p.590).

$$\mathrm{t}_{\mathrm{test}} = \frac{b_i}{S_{b_i}} \qquad \textbf{Equation 2-3}$$

The regression model described in Equation 2-2 only considers first-order terms, also referred to as main effects. The use of main effects is ideal for screening and can produce accurate models. In some cases the effect of an independent variable changes due to the influences from another independent variable. These interactions can be modelled by using an interaction term that describes the combined interaction between the independent variables. Equation 2-4 shows the interaction term for a two-way interaction where X_3 describes the influences of the interaction between the two independent variables X_1 and X_2. The final interaction equation for two independent variables is shown in Equation 2-5 (Berenson et al. 2012, p.602).

$$X_3 = X_1 \times X_2 \qquad \textbf{Equation 2-4}$$

$$\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_1X_2 = b_0 + b_1X_1 + b_2X_2 + b_3X_3 \qquad \textbf{Equation 2-5}$$

Interaction terms can also be extended to three-way interactions and higher but this is rarely done in practice (Devore & Farnum 2005, p.519; Ziyani & Fatah 2014). Curvilinear regression models are another form of regression models that uses polynomial equations. The quadratic equation is one of the most common polynomials used. Transformations can also be used to transform the data to another scale that can then be used to develop MLR models. Logarithmic, square-root or reciprocal transformation are some of the transformations that can be used for MLR modelling (Berenson & Levine 1989, p.664; Devore & Farnum 2005, p.521).

2.2.3 Artificial neural networks

Feed forward neural networks are widely applied in many applications such as forecasting. An ANN is a very powerful nonlinear technique that can be used in almost every situation where there exist a relationship between independent and dependent variables (Azadi & Karimi-Jashni 2016; Hill & Lewicki 2006, p.419).

The technique was formed through research attempts to model the brain's capacity to learn. Neurons receive inputs that are weighted to correspond to the strength of the signal. The weighted sum from the inputs is used to determine the activation and the signal is passed through to produce an output (Hill & Lewicki 2006, p.420).

Feed forward neural networks consist of an input, output and some hidden layers. The structure is referred to as the topography of a neural network. The number of hidden layers and neurons per hidden layer can vary according to the dimensions of the problem at hand. There exist some rules-of-thumb that can assist in obtaining the optimal topography but in

most cases it is achieved through trial and error. Feed forward neural networks are trained in two phases wherein the neural network is firstly fully calculated and then secondly the adaption of the weights from the prediction error (Abraham 2004; Hill & Lewicki 2006, pp.421–422; Satish & Setty 2005).

The first phase is referred to as the forward pass where data from the independent variables are introduced through the input layer of the neural network. The number of input neurons is equal to the number of independent variables. The input layer is connected to an adjacent hidden layer where each connection has its own weight, representing the strength of the connection. The input value at the new node is calculated by taking the sum of the product between the values of all the input neurons (Xi) and its weight connections (W_{ij}), where j represents the layer number, i the input neurons and o the output neurons in layer j. The input value passes through an activation function (f) to determine the output value for the node. This process is repeated until all the neurons are calculated. Equation 2-6 is used to calculate the neuron values. Each layer has a bias neuron ($bias_j$) with a value of one to serve as a constant input (Azadi & Karimi-Jashni 2016; Hill & Lewicki 2006, p.422; Satish & Setty 2005).

$$Node\ value = f\left(\sum_{i=1}^{m} W_{i0j}X_i + bias_j\right) \qquad \textbf{Equation 2-6}$$

Different activation functions can be used to achieve different results. Some have mathematical limitations to the output values. Different activation functions can influence the training process due to the derivative of the activation function being used in most training algorithms. Some of the most common activation functions include the pure-linear, log-sigmoid and tan-sigmoid functions as shown in Figure 2-6 (Ayat & Pour 2014; Jain et al. 1996; Topuz 2010).

The second phase is called the back pass that only occurs during training where a training algorithm is used to adjust the weights of the neural network starting at the outputs neurons and progressing backwards towards the input neurons. The training algorithm is used to minimize the error of the ANN by adjusting the weights of the connections (Ghaffari et al. 2006).

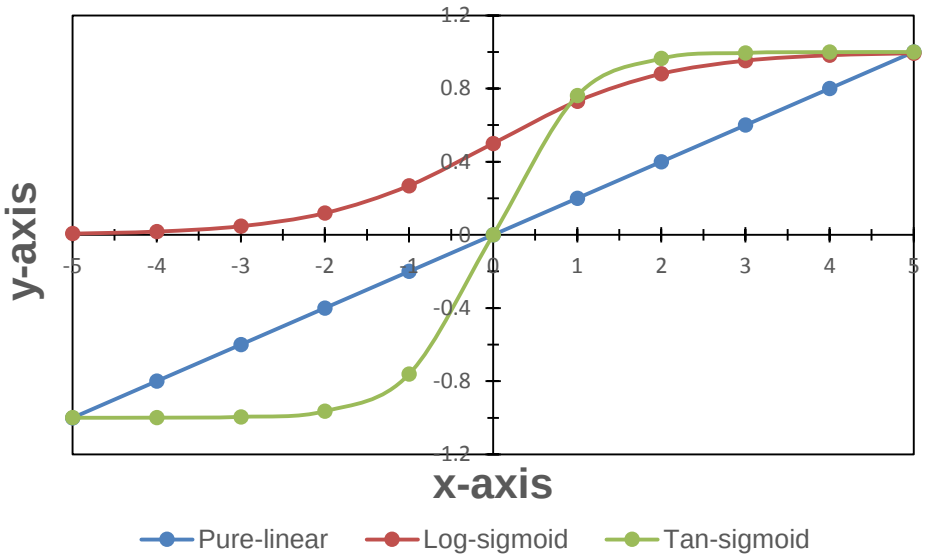


Figure 2-6: Common activation functions for artificial neural networks adapted from Ayat & Pour (2014)

ANNs are trained to minimise the error of the training dataset that might lead to overtraining. ANNs with too many weights are generally prone to over train resulting in an over-fitting of the modelling data due to the complexity of the model. A more complex topography will also influence the performance of an ANN. Over-fitting is not ideal as it causes the model to lose its ability to generalise. The model fits the modelling data points to a fair extend but struggles to predict data points that are in-between. Too few weights on the other hand may result in an insufficiently constructed model that can experience undertraining. Using validation data as an early training stopping technique improves the training process and prevents overtraining. The validation error is checked to determine whether a further increase in training will result in a decrease in the error. Overtraining occurs when the error of the validation data increases while the training error decreases, resulting in the neural network over-fitting the training data and miss-predicting the validation data (Hill & Lewicki 2006, pp.430–432).

The error valley problem is another common problem that could exist. There may exist different local minima's on the error surface in which the algorithm can become stuck in as shown in Figure 2-7. The neural network can miss the global minimum and become less accurate (Kriesel 2007, p.53; Wilamowski & Irwin 2011, p.1).

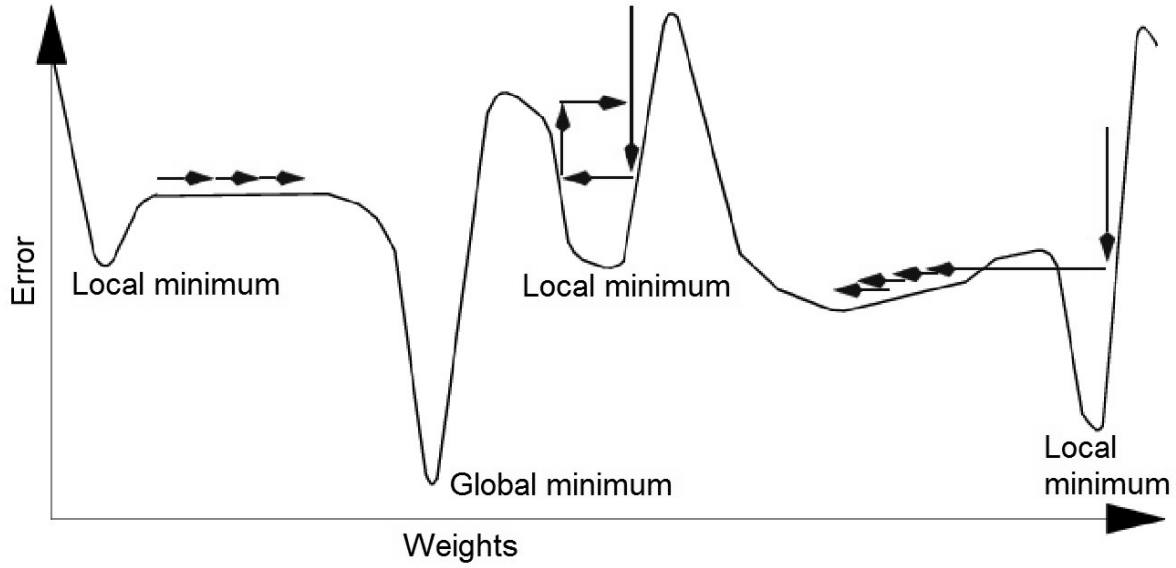


Figure 2-7: The error-valley problem adapted from Kriesel (2007, p.53)

The gradient descend training algorithm is one of the most common training algorithms used. This algorithm takes small step changes that make it stable but with a slow convergence which can fall into the error valley problem. The Gauss-Newton training algorithm can overcome the error valley problem by taking the second-order derivative of the error function to evaluate the curvature of the error surface. The Levenberg-Marquardt algorithm is a combination of both these training algorithms, with the stability and convergence of the gradient descend algorithm and the speed and strength of the Gauss-Newton algorithm (Abraham 2004; Wilamowski & Irwin 2011, pp.1–2).

The Levenberg-Marquardt algorithm uses the Jacobian and Hessian matrices that are the first and second order derivative matrices of the error function. The weights are updated using Equation 2-7 where *K* represents the iteration. The weight matrix (***W***) is an *N* x 1 matrix where N is the total number of weights. The Jacobian matrix (***J***) is a (*n* x *k*) x *N* matrix, where *n* represents the training samples size and *k* the total amount of independent, output, variables. The identity matrix (***I***) is an *N* x *N* matrix and the error matrix (***e***) is an (*n* x *k*) x 1 matrix. A small combination coefficient (μ) will result in the weight updating algorithm to approximate a gradient descend algorithm whereas a large training rate will approximate the Gauss-Newton algorithm (Abraham 2004; Wilamowski & Irwin 2011, p.7).

$$\boldsymbol{W}_{K+1} = \boldsymbol{W}_K - (\boldsymbol{J}_K^T \boldsymbol{J}_K + \mu \boldsymbol{I})^{-1} \boldsymbol{J}_K^T \boldsymbol{e}_K \qquad \textbf{Equation 2-7}$$

2.3 Critical literature review

The FBG process is a complex process with different factors influencing the growth mechanisms. These growth mechanisms results in granules with different properties. The modelling of the process plays a key role in determining the variables that have the largest influence on the granulation process. Inferential models can be developed by investigating the correlations between the operating, input, variables and the output variables. These inference models can be used to control the granulation process to the desired output.

Most of the authors investigated the operating variables, but some of the operating variables, i.e. SLT and SSL variables, have not received much attention. Most of the correlation results were obtained from regression analysis and not by means of an additional statistical technique. Only Ziyani & Fatah (2014) studied two-way interaction effects MLR models. Interactions are expected due to the highly integrated process and further studies are required. Most of the authors who investigated MLR models only reported the CMD with no error values. High CMD values can be obtained even when the models do not fit the data accurately as discussed in section 2.2.1.

Chapter 3. Experimental Procedures

The influence and relationship of the operating variables on the output variables are evaluated by conducting experiments on a continuous industrial fluidised bed granulator (FBG). Different samples are taken and analysed to obtain the values for the output variables. A screening phase, phase one, is introduced to determine the significance of the operating variables. A larger number of experimental runs are conducted for model development in phase two. The multiple linear regression (MLR) modelling technique are qualitatively compared and evaluated with the artificial neural network (ANN) modelling technique using different statistical parameters. All the models will be assessed to determine the performance and usability.

3.1 Experimental setup

An industrial Prill Granule Ammonium Nitrate (AN) fluidised bed granulator is used to conduct experimental trials. AN granules are used as seed particles which are sprayed with an liquid AN and a Galoryl AT plus solution. Different operating and output variables are measured on-site and logged to be used in model development. The feed, granulator product and final product streams are sampled and then analysed at the on-site laboratory to obtain the size distribution, circularity and porosity.

3.1.1 Materials used

Dry AN granules are used as seed material with an average mean particle size of 1.2mm. A spray solution of liquid AN, water and Galoryl AT plus binder is used. Different spray concentrations are achieved by varying the amount of water. The average AN concentration in the spray solution is 92%. Atmospheric air is used as process air for fluidisation and atomisation. All the materials used in the granulation process is provided by the plant and summarised in Table 3-1.

Table 3-1: Materials used

Material	Description/use
Dry ammonium nitrate	Seed particles
Liquid ammonium nitrate	Spray liquid
Water	Solvent in spray liquid
Galoryl AT plus	Binder in spray liquid
Air	Process air for fluidisation and atomisation

3.1.2 Process flow diagram

AN seed particles are fed with a screw feeder at point A to the FBG as shown in Figure 3-1. Fluidising air is introduced into the granulator via a fan. The air is heated by a heat exchanger at point B and enters the granulator at points C and D. The air is used to fluidise and dry the particles inside the granulator. A solution of AN melt, water and Galoryl AT plus binder is sprayed onto the particles using nozzles at point E. The final granulator product exits the granulator at point F and is transported with a conveyor belt to a set of screens at points G and H where the over- and undersize particles are separated from the on-size particles. The over- and undersize particles are recycled back to the process at points I and J. The oversize particles are crushed with a jaw crusher at point K and are subsequently fed along with the undersize particles to the granulator. The final product is transported on a conveyor belt to a storage facility at point L. Blow-off fines are recovered in a wet scrubber where it is recycled as molten AN.

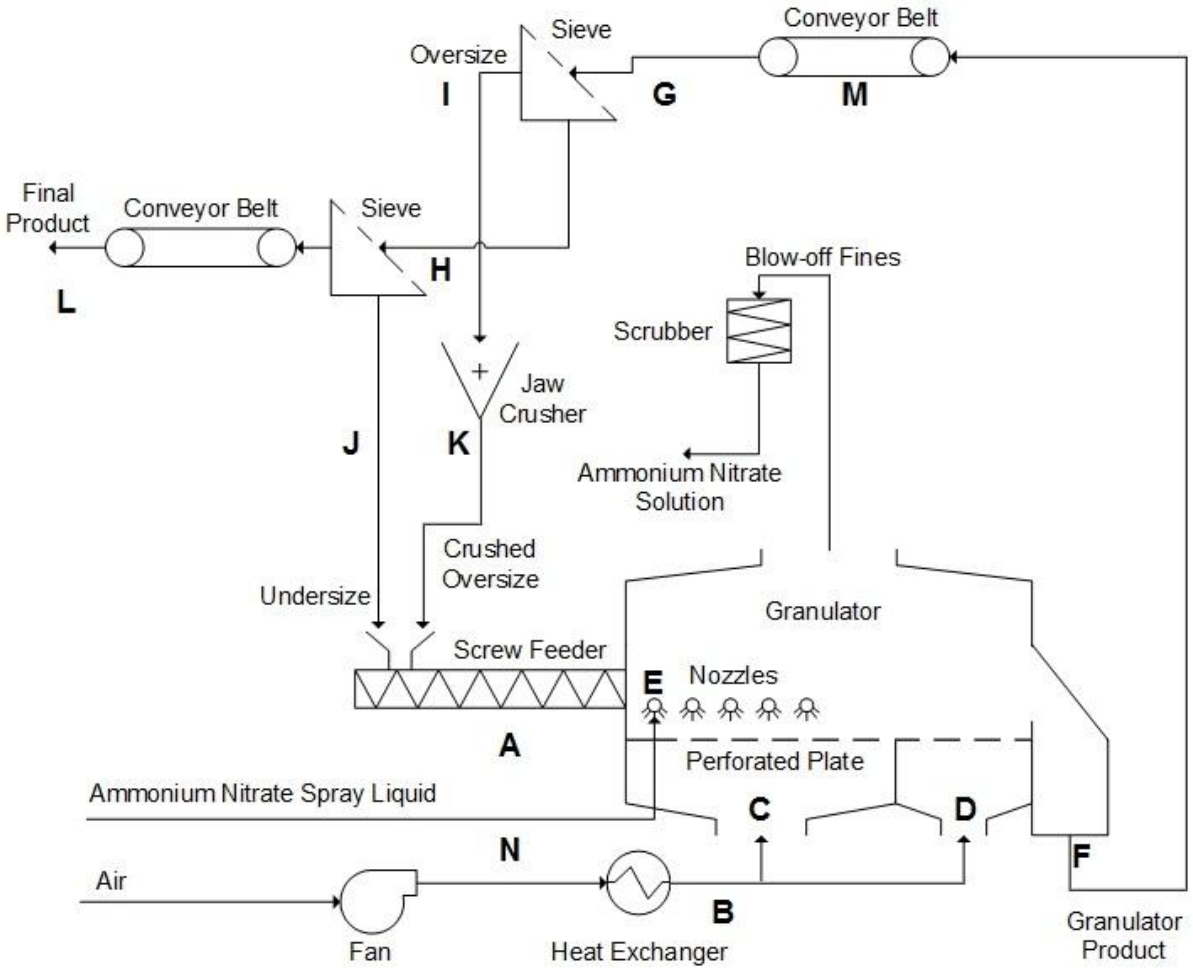


Figure 3-1: Process flow diagram with points of interest

3.1.3 Process sampling

The feed is sampled via a sampling point underneath the crusher at point K on Figure 3-1. This was done as the influence of the crusher was the main contributor to the feed size and restrictions prevented sampling of stream J. These samples were used to obtain the feed particle size distribution (PSD) where the mean particle diameter (d_{50}) and the particle distribution slope values are calculated. The influence of the crusher on the seed particles is measured by sampling the feed.

The granulator product is sampled from the conveyor belt by taking a cut of the belt at point M. A PSD is obtained for the granulator product where it is used to evaluate the influences of the operating variables on the size properties of the particles exiting the granulator. The granulator product stream is sieved and the oversize granules are crushed and recycled along with the fines. The on-size particles are transported to a storage facility and are referred to as the final product.

The final product of the granulator is sampled from the conveyor belt by taking a cut of the belt at point L. These samples are used to obtain the final product PSD, porosity and circularity. The final product consists of the granulator product that is screened to a specific size range, i.e. 1.6mm and 3.0mm.

3.1.4 Operating variables

The input and output variables of the FBG that were investigated for this project are summarised in Table 3-2.

Table 3-2: Summary of the process variables

Input variables	Output variables
Fluidised air flow rate (FAF)	Production rate (RUR)
Fluidised air temperature (FAT)	Recycle rate (RER)
Spray liquid flow rate (SLF)	Efficiency (EFF)
Spray liquid temperature (SLT)	Product mean particle size (PPS)
Spray liquid concentration (SLC)	Granulator mean particle size (GPS)
Seed mean particle size (SPS)	Product porosity (POR)
Seed (linearized PSD) particle slope (SSL)	Product circularity (PCR)
	Product (linearized PSD)particle slope (PSL)
	Granulator (linearized PSD)particle slope (GSL)

The fluidising air flow rate (FAF) is adjusted using the air damper on the cooling air side at point D as shown on Figure 3-1. Closing the damper will result in an increase in the fluidising air flow rate at point C. The FAF is measured using a Yokagawa venturi flow meter at point C. The fluidising air temperature (FAT) is adjusted using an air heat exchanger at point B to increase the temperature of the incoming air and is measure by a WIKA type K thermocouple at point B.

The number of nozzles is adjusted at point E by opening or closing nozzles to achieve different spray liquid flow rates (SLF). The flow rate is measured using a Micromotion Coriolis mass flow meter at point N. The spray liquid temperature (SLT) is controlled by the temperature inside the collection vessel located upstream from the FBG. A heating jacking around the collection vessel is used to regulate the temperature. The SLT and collection vessel temperatures are measured using WIKA type K thermocouples. Both the SLF and SLT are measured at point N.

The concentration of the ammonium nitrate (SLC) is controlled by changing the evaporator temperature and pressure. The evaporator is upstream from the FBG in stream N. The temperature and pressure of the evaporator are measured using a WIKA type K thermocouple and Rosemount Chemiseal pressure transmitter. The AN concentration is calculated using a plant-proprietary material properties spreadsheet together with the measured temperatures and pressures from the collection vessel. The evaporator pressure along with a guessed concentration value is used to calculate the evaporator temperature from the spreadsheet. The correct concentration value is obtained by adjusting the guessed concentration value until the calculated evaporator temperature is the same as the measured evaporator temperature.

The seed mean particle size (SPS) is adjusted by adjusting the number of spacers on the crusher at point K. A sample is taken underneath the crusher at point K on Figure 3-1 each time to ensure that the seed particle size is adjusted according to the experimental plan.

All the electronic measurements are logged using a Delta V DCS and the Osisoft PI historian software package. The values are logged in minute averages to reduce variance in the data.

3.1.5 Definition of output variables

3.1.5.1 Production rate

The production rate (RUR), or production rate, is the amount of final product produced per hour. It is calculated from a steady state mass balance over the whole granulation plant using the assumption that the amount of final product produced is equal to the amount of fresh ammonium nitrate ($AN_{Feed\ to\ FBG}$) entering the FBG. The production rate is calculated using Equation 3-1.

$$\mathrm{RUR} = \mathrm{AN}_{\mathrm{Feed\ to\ FBG}} \qquad \textbf{Equation 3-1}$$

3.1.5.2 Recycle ratio

The recycle ratio (RER) is the ratio of the amount of product recycled from the total amount of product exiting the FBG. The recycle ratio is calculated from a mass balance between the total production rate, the total amount of product produced per hour, and production rate using Equation 3-2.

$$\mathrm{RER} = \frac{\mathrm{Total\ production\ rate} - \mathrm{RUR}}{\mathrm{Total\ production\ rate}} \qquad \textbf{Equation 3-2}$$

3.1.5.3 Efficiency

The efficiency (EFF) of the plant is the ratio of the amount of fresh AN entering the FBG to the amount of AN sprayed. It is calculated from a steady state mass balance with the assumption that the AN entering the plant has a concentration of 91%. The SLF value is always higher than the fresh feed because a certain amount of spray liquid will be recycled back. The correct spray liquid concentration (SLC) is used to adjust the efficiency value for concentration. The efficiency can therefore be calculated using Equation 3-3.

$$\mathrm{EFF} = \frac{\mathrm{AN}_{\mathrm{Feed\ to\ FBG}}}{\mathrm{SLF}} \times \frac{\mathrm{SLC}}{91\%} \qquad \textbf{Equation 3-3}$$

3.1.5.4 Particle size distribution and circularity

The PSD and circularity are calculated for the crusher, granulator and product samples using a Haver CPA 2-1 laser diffraction particle size analyser. Scattered light from the particle is detected by a camera which is used to calculate the physical properties of each particle (Napier-Munn & Wills 2006, pp.104–105). This method of particle size analysis gives fast, reproducible results within a wide size range (Napier-Munn & Wills 2006, p.105; Rhodes 2008, p.16).

The mean particle size is used as a representative measurement for the particle sizes of the different streams. The d_{50} value represents the size that equally splits the distribution into half (Rhodes 2008, pp.7–8).

The slope of the PSD represents the sharpness of the distribution. A larger slope will result in a larger number of particles within a specific range, a sharper distribution, whereas a lower slope will result in particles that are more evenly distributed as shown in Figure 3-2 and Figure 3-3 (Macias-Garcia et al. 2004).

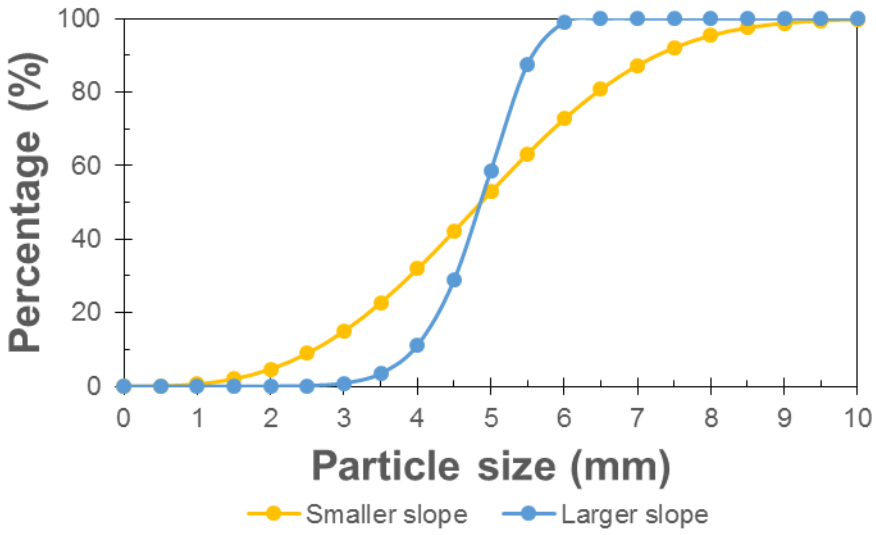


Figure 3-2: Cumulative particle size distributions with different slopes adapted from Rhodes (2008, p.8)

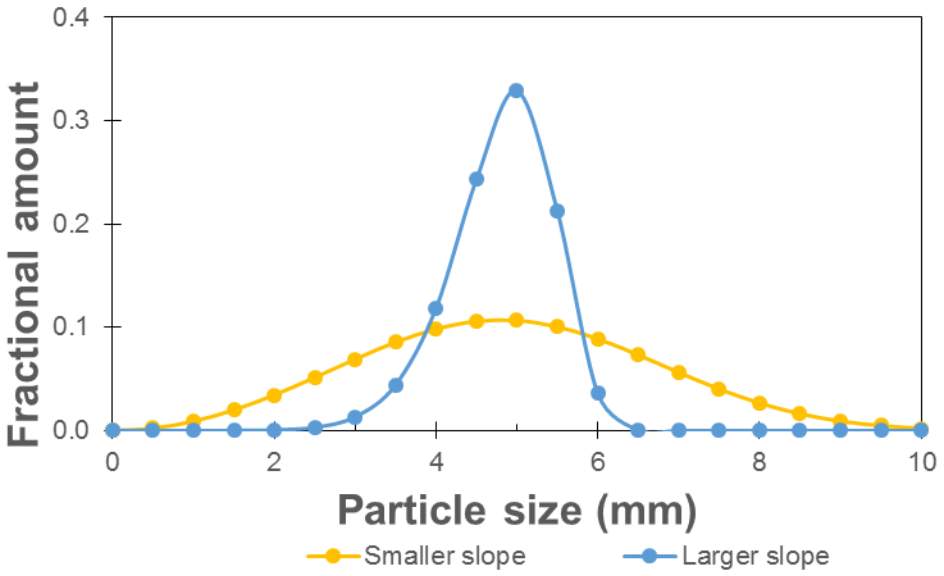


Figure 3-3: Differential frequency distribution with different slopes adapted from Rhodes (2008, p.4)

The slope (*m*) is calculated using the first quartile (d_{25}) and third quartile (d_{75}) values in the linearized Rosin-Rammler equation. The full linearization is found in Appendix A. The slope of the distribution can be calculated using Equation 3-4 with (Macias-Garcia et al. 2004).

$$m = \ln\left(\frac{\ln[1-0.25]}{\ln[1-0.75]}\right) \div \ln\left(\frac{d_{25}}{d_{75}}\right) \qquad \textbf{Equation 3-4}$$

Circularity is defined as the amount to which the perimeter of a particle (*P*) is related to a circle. Circularity is calculated by the ratio of a particle's projected area (*A*) to the perimeter of the particle (Pashminehazar et al. 2016). A circularity of 1 indicates that the projected area of the particle (r_A) and the perimeter (r_P) of the particle have identical radiuses, just like a circle. Equation 3-5 is used to calculate the circularity of a particle (Aleksic et al. 2014).

$$\text{Circularity} = \frac{4\pi A}{P^2}\left(= \frac{4\pi^2 r_A^2}{4\pi^2 r_p^2} = \frac{r_A}{r_P}\right) \qquad \textbf{Equation 3-5}$$

3.1.5.5 Porosity

The porosity variable is used to ensure the absorption quality of the final product. Two different approaches, liquid absorption and computerised tomography, were investigated to determine the similarity between the techniques and to verify whether the cheaper, liquid absorption, technique could be used. The liquid absorption technique exposes the sample to a liquid where the volume of the absorbed liquid is equal to the pore volume of the sample (Maria 2010; Webb 2001). The sample is weighed before and after the process to calculate the weight of the absorbed liquid. The densities are used to convert the mass values to volume values. These values are used in Equation 3-6 to calculate the volume porosity. Diesel with red food colouring was used since AN is insoluble in organic solvents.

$$\text{Porosity} = \frac{\text{Volume of voids}}{\text{Volume of solids} + \text{Volume of voids}} \qquad \textbf{Equation 3-6}$$

A fixed amount of sample is used to ensure consistency and comparable results. The oil level is started at the same level to ensure that the hydraulic pressure is equal for all analysis. Sufficient time is allowed for oil to absorb according to a set standard in order to yield repeatable measurements. The oil absorption setup can be seen in Figure 3-4.

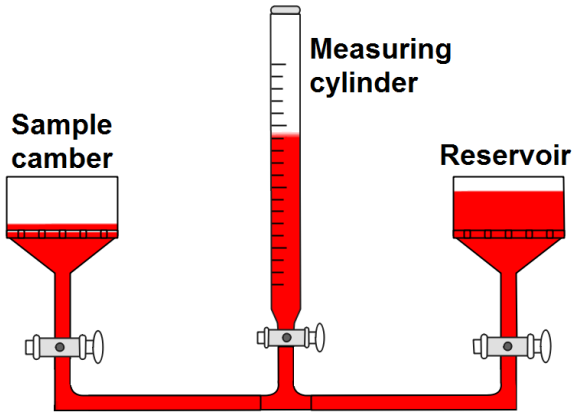


Figure 3-4: Oil absorption setup

Computerised tomography (CT) provides a quantitative detection of the interior structure of a sample without altering or destroying the sample. The technique uses a set of x-rays to obtain a two-dimensional image of the sample. X-rays pass through the open or less dense areas effortlessly and are received by a detector. The x-rays that encounter dense areas do not pass through easily, resulting in lower energy x-rays received by the detector. The difference in the energy levels of the x-rays results in the image. At least 180° of the sample needs to be scanned in order to for this technique to produce accurate results (Pashminehazar et al. 2016; Yao et al. 2009).

For the purpose of this project 1000 slices were scanned at a 360° rotation with a 225 kV x-ray source. Three-dimensional models were constructed from the images and underwent different image processing techniques in order to calculate the volume of the voids and solids.

The results from 58 samples the liquid absorption and CT scan techniques were statistically evaluated using a dependent t-test. Table 3-3 summarises the results between the techniques.

Table 3-3: Statistical results from porosity methods

	Oil absorption technique	CT technique
Mean	35.30%	29.16%
Standard deviation	5.16	3.92
Coefficient of variance	14.6%	13.4%

The CT technique shows the lowest standard deviation with a 1.24 difference. The standard deviation parameter is used to measure the deviation of the data from the mean (Field 2009, p.42). It is difficult to draw any conclusions from the standard deviation values if two dependent variables have the same units but large differences in the mean values. The coefficient of variation (CV) measures the deviation of the data (σ) relative to the mean ($\bar{X}$). The CV parameter is useful when comparing datasets with large deviations in the mean values and standard deviations. The CV value is calculated using the standard deviation and mean in Equation 3-7 (Berenson & Levine 1989, pp.50–51). The CV values for both techniques show a small difference of 0.8% indicating that the data from both techniques have a near equal deviation from their means.

$$\mathrm{CV} = \frac{\sigma}{\bar{X}} \times 100 \qquad \textbf{Equation 3-7}$$

A p-value (two-tailed) < .05 is obtained for the pair test, indicating a close relation between the two techniques. Statistical significance tests are unreliable to infer the importance of the results because of the dependency of the p-values on the dataset size. Larger datasets will result in a decrease in the p-values, with all other factors being constant, making it difficult to assess the significance of the results. The effect size is independent of the dataset size and can be used to describe the significance. Cohen (1992) interpreted the effect size as a small effect for values around 0.2 and medium effect for values near and around 0.5 with a large effect for values higher than 0.8. The effect size is calculated using the means from both data sets ($\bar{X}_1$ and $\bar{X}_2$) and the maximum value between their respected standard deviations (s_1 and s_2) as shown in Equation 3-8 (Ellis & Steyn 2003).

$$\text{Effect size} = \frac{|\bar{X}_1 - \bar{X}_2|}{\max(s_1, s_2)}$$ **Equation 3-8**

A high effect size of 1.19 was obtained indicating that there is a large difference in the values obtained between the two techniques. The oil absorption technique has a larger mean and standard error compared to the CT technique. The large differences resulted in the high effect size value.

The high mean difference, low difference in the CV values and large effect size indicates that there is a constant offset between the two techniques as shown in Figure 3-5, apart from a few points. A constant offset will not influence the results from the correlation values between the operating variables and the porosity variable. The coefficients for the independent variables will differ but can only be adjusted.

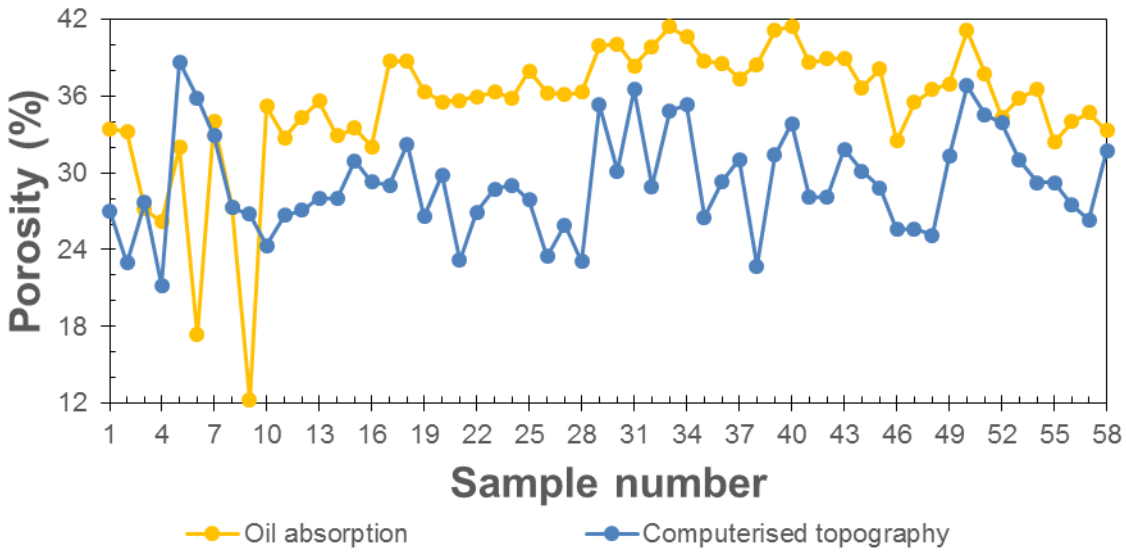


Figure 3-5: Porosity results from both techniques

The oil absorption technique is chosen to determine the porosity as it is the most cost-effective and fastest method that is daily used for plant operations.

3.2 Experimental procedure

The experimental procedure for this study is divided into two phases due to the high number of variables and complexity of the process. Phase one is considered as a screening phase and familiarisation of the process equipment and plant operation. Phase two consist of the bulk of the experiments which are used in the development of the final models. Validation is the final stage in the modelling phase and is required to assess the accuracy and strength of the developed models.

Randomised experimental designs are used to eliminate bias towards the experimental results, ensure that a representative sample of the population is obtained and helps to distribute the unknown variation in the process. There are different randomised experimental design approaches, with the most common being the complete randomised design and the randomised block design (Hinkelmann Klaus 1994, p.145; Oehlert 2010, pp.315–316). The complete randomised design assigns values to independent variables completely randomly (Jayaraman 2000, p.70). Randomised block designs eliminates a known variation source to lower the experimental error (Jayaraman 2000, p.79). The variables are grouped into blocks and then randomly assigned in order to block the influence (Montgomery 2008, pp.139–140). The complete randomised experimental design is followed in this study to assess all the influences from the operating variables. Randomised replication is also used to calculate the error variance of the experiments (Jayaraman 2000, p.69).

3.2.1 Phase one

The first phase is a screening phase where only main effects between the variables are evaluated (MLR-M). Focussing on only the main effects lowers the number of runs required. Multiple linear regression (MLR) is chosen as the modelling technique since the data show a monotonic behaviour.

Three operating levels are used for each variable (except for the spray liquid concentration) and these scenarios are executed according to the complete randomised experimental design. This results in scenarios where all the variables vary from the previous run. The experimental plan along with the ranges for the variables can be found in Appendix B.

The randomised approach of the experimental plan allowed for multiple operating parameter changes after each run. The results from the first phase are used to qualitatively evaluate the effects of the operating variables on the output variables. A variable with no effect is discarded, leaving only the statistically significant variables for investigation.

3.2.2 Phase two

Main effects (MLR-M) and interaction effects (MLR-I) multiple linear regression models are developed in the modelling phase. The inclusion of interaction effects are evaluated by comparing both models. Including interaction effects can increase the accuracy and performances of the MLR models.

Five different operating levels are used to increase the distribution of data points. The experimental plan for phase two is also executed using the complete randomise experimental design where the values from each level are randomly assigned for each run. The experimental plan along with the values for each level can be found in Appendix B.

The spray liquid concentration could only be divided into three levels. The spray liquid temperature had different ranges for each concentration level based on operational limitations. The values for each of the temperature levels are calculated for each concentration level and are summarised in Table B-5 in Appendix B.

Some experiments are repeated in order to determine the experimental error using t-confidence intervals (Devore & Farnum 2005, pp.313–319). A 95% confidence interval is used to calculate the error as seen in Appendix A. Relative small error percentages are obtained, with the production rate and product porosity having the highest average percentage error of 16%, followed by the PSD slope variables with errors around 12%. Table 3-4 summarises the results for the repeatability of the experiments.

Table 3-4: Repeatability results for phase 2

Quality variable	Mean	Standard deviation	Error	Average percentage error
Production rate	3.18	0.41	0.51	16%
Recycle ratio	8.09	0.52	0.65	8%
Efficiency	0.76	0.04	0.05	6%
Product mean size	1.98	0.11	0.14	7%
Granulator mean size	1.55	0.04	0.05	3%
Product porosity	33.31	4.19	5.21	16%
Product circularity	0.88	$3x10^{-3}$	$4x10^{-3}$	0.5%
Product PSD slope	5.50	0.59	0.73	13%
Granulator PSD slope	4.49	0.45	0.55	12%

The MLR models are subsequently compared with a non-linear technique, artificial neural network (ANN). Models by both techniques are developed using the exact same data in order to compare the performances between both modelling techniques. The dataset is divided into

80% for model development and 20% for validation (Hyndman 2014, p.51). Validation is necessary to evaluate a model's performance on unseen data.

3.3 Model development

The MLR modelling technique is chosen to model the influences between the operating variables and output variables. Main effects and interaction effects MLR models were developed using IBM SPSS Statistics for Windows, Version 23.

ANN models were chosen as a comparative modelling technique for the MLR models. The ANN models were not optimised as rule-of-thumb topographies were chosen and were only used, in rudimentary form, as a qualitatively comparison. The Neural Network Toolbox™ of MATLAB® was used with most of the parameters left to their default values. Both modelling techniques were evaluated and compare using different statistical parameters.

3.3.1 Spearman's rho correlation

Spearman's correlation coefficient is also referred to as Spearman's rho value that is a non-parametric statistic that can be used regardless of the data distribution. The Spearman's rho (ρ) values are used to determine the strength of the relationship between variables. Spearman's coefficient does not measure the linearity of the relationship but assess the degree to which a monotonic function can describes the relationship between two variables. A function that only increases or decreases is referred to as a monotonic function as shown in Figure 3-6 (Field 2009, pp.179–180; Hauke & Kossowski 2011; Legendre & Legendre 1998, pp.197–198).

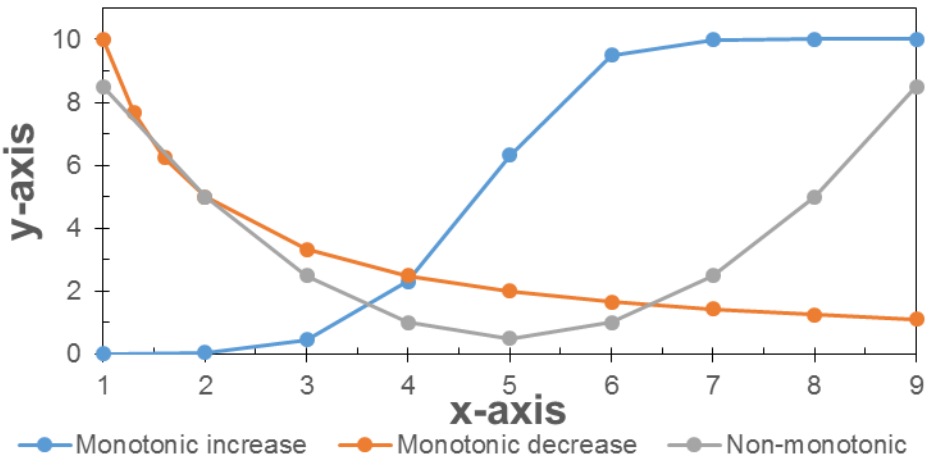


Figure 3-6: Monotonic and non-monotonic functions adapted from Hauke & Kossowski (2011)

Covariance (cov_{xy}) is used to determine the relationship between two variables. It is dependent upon the scale of the measurements and requires to be standardised with the standard deviation. The Spearman's correlation coefficients are calculated by firstly ranking the data from high to low and then using Pearson's correlation equation given by Equation 3-9 with *n* representing the sample size (Berenson & Levine 1989, pp.542–543; Field 2009, p.180).

$$\rho = \frac{cov_{xy}}{s_x s_y} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)s_x s_y} \qquad \textbf{Equation 3-9}$$

The correlation coefficient produces values between one and negative one. Values close to one indicates a strong positive relationship between the variables whereas a value close to negative one indicates a strong negative correlation. A value of zero indicates that there exists no relationship and the two variables are independent. It is important to note that an increase in one variable might result in an increase in another but not necessarily the other way around. It is therefore necessary to evaluate and assess the results of the Spearman coefficients. The statistical significance of the rho values must also be tested using a hypotheses test, which can be directional or non-directional. A directional hypotheses test is used when it is known whether the value will decreases or increases and is tested using a one-tailed test. A non-directional hypothesis is used when the value can either increase or decrease and is tested with a two-tailed test. The correlation is regarded as statistical significant if the p-values from the test have a value less than .05, this indicates a 95% certainty (Berenson & Levine 1989, p.549; Field 2009, p. 54,181).

3.3.2 Multiple linear regression model development

Selecting the independent variables is a very important role in the development of MLR models. There are three different methods referred to as the block wise entry, force entry and stepwise entry. The block wise entry method starts with adding independent variables found from previous studies and then adding new variables one at a time. The forced entry method adds all the independent variables simultaneously and is favoured by most researches since it is independent from influences from literature. The stepwise method adds independent variables based on a mathematical criterion. The statistical models in this study are developed using the forced entry method to reduce influences from literature and random variation from the data (Field 2009, p.212).

A t-test is used to determine the statistical significance of each variable, where the p-values are used to evaluate the significance of the operating variables with regards to the quality

variable. A p-value less than .05 would be classified as significant, resulting in models having a 95% confidence (Field 2009, p.228). Only significant variables are included in the final models.

The Shapiro-Wilk test of normality is used to evaluate whether the data are normally distributed. The Shapiro-Wilk test first assumes that the residuals are normally distrusted. The residuals from the data are compared with residuals of a normal distribution in order to test the assumption. The test will be non-significant if the residuals are normally distributed resulting in a p-value larger than .05. It therefore accepts the normality hypothesis and is classified as normally distributed (Azadi & Karimi-Jashni 2016; Field 2009, p.144).

The data are also evaluated for any outliers using standardized residual box plots. Standardized residuals are the residuals between the predicted and actual values that are standardized using the standard deviation. Using standardized residuals gives the capability to compare the residuals from different models. Any definite outliers are identified from the box-plot, evaluated and removed from the dataset and new models are developed (Field 2009, p.216).

Using interaction effects increases the number of independent variables for modelling. Table 3-5 lists all the possible interaction effects for the chosen independent variables.

Table 3-5: List of interaction combinations

Interaction combinations		
FAF*FAT	FAT*SLT	SLF*SSL
FAF*SLF	FAT*SLC	SLT*SLC
FAF*SLT	FAT*SPS	SLT*SPS
FAF*SLC	FAT*SSL	SLT*SSL
FAF*SPS	SLF*SLT	SLC*SPS
FAF*SSL	SLF*SLC	SLC*SSL
FAT*SLF	SLF*SPS	SPS*SSL

3.3.3 Artificial Neural Network model development

ANN models were developed and used as comparison to the statistical MLR models. The ANN models are trained with the exact same dataset that is used to develop the statistical models. The dataset is divided into 80% for modelling and 20% for validation.

The data for the ANN models is divided index wise with the first index consisting of the same training data as the statistical models, the second index contained the validation dataset. The default data processing steps are used, where the data are normalised between -1 and 1.

Normalisation is done to ensure that the sigmoid function does not become saturated and only give out the extreme values of -1 and 1. If this occur, especially at the beginning, it will result in a slow step change and therefore a slow training phase. Another processing step includes dividing the data into training and validation sets (Beale et al. 2014, p.63).

Literature suggests that one hidden layer is sufficient to approximate most complex nonlinear functions (Azadi & Karimi-Jashni 2016). The number of hidden neurons in the hidden layer is calculated using Equation 3-10 (Karsoliya 2012; Sharma et al. 2013).

$$\text{Number}_{\text{Hidden neurons}} = \frac{2}{3} \times \text{Number}_{\text{Input}} + \text{Number}_{\text{Outputs}} \qquad \textbf{Equation 3-10}$$

The tan-sigmoid and pure-linear activation functions are used in the hidden neurons and output neurons respectively. The activation functions are given by Equation 3-11 and Equation 3-12 with *x* representing the function input value. All the initial weights are set to a constant value of 0.5 in order to obtain consistent results for evaluation and comparison.

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \qquad \textbf{Equation 3-11}$$

$$f(x) = x \qquad \textbf{Equation 3-12}$$

The Levenberg-Marquardt training algorithm is used with all the training parameters set to their default values as summarised in Table 3-6. The validation dataset is used to prevent overtraining by calculating the error between the prediction and actual outputs after each epoch. Training is stopped should the validation error fail to decrease beyond the maximum for 10 consecutive epochs and the results from the best case are used. The mean square error (MSE) is used to evaluate the performance of the ANN models calculated by Equation 3-13 (Beale et al. 1992, pp.4–8).

Table 3-6: Neural network training parameter values Beale et al. (2014, p.327)

Neural network training parameter	Value
Maximum number of epochs	1000
Performance goal	0
Maximum validation failures	10
Minimum performance gradient	$1x10^{-7}$
Initial mu	0.001
Mu decrease factor	0.1
Mu increase factor	1
Maximum mu	$1x10^{10}$
Initial weight value	0.5

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2$$ **Equation 3-13**

The maximum number of epochs represents the maximum number of iterations that the ANN is allowed to complete before stopping. Successful training is usually achieved before the maximum number of epochs is reached, supporting the decision to leave it to its defaults value. The performance goal is the minimum error value set for the ANN to reach, with zero is usually the desired value. The minimum performance gradient helps to ensure that a change in the weight value is taken and that the training phase does not become stagnant. Since small changes need to be taken to ensure that the global minimum is reach, the value was left to its default. The mu variable is the combination coefficient which is inversely proportional to the training rate. It therefore determines the magnitude of the change taken in each weight value during each training iteration and is adjusted according to the increase or decrease factor depending on whether an increasing or decreasing change was made. The default values were chosen since there was no information that could indicate whether a fast or slow training can be expected (Beale et al. 1992, pp.138, 157).

Two sets of ANN models were developed. The first consists of the operating variables that were found to be statistical significant in the main effects MLR models (ANN-M), while the second consists of all the operating variables as inputs (ANN-I). The first set evaluates the accuracy of ANN models developed using the results from the main effects MLR models (MLR-M), while the second ANN models assess the performance of ANN using all the independent variables.

3.3.4 Model performance evaluation

There is no universally applicable parameter to measure the accuracy and performance of different models. Different statistical parameters are chosen to measure the accuracy and performance of each model. Each parameter has its own advantages and disadvantages, but using different parameters increases the accuracy of the performance evaluation (Azadi & Karimi-Jashni 2016).

Models can easily over fit the training data and can give a false estimation of the accuracy of the model. The accuracy and performance of models can accurately be evaluated using a set of validation data, which was unknown to the model. The standard error of estimate (SEE), (S_{YX}), and coefficient of multiple determination (CMD), (R^2), is useful in evaluating the model's performance while the mean absolute error (MAE), root mean square error (RMSE) and the

mean absolute percentage error (MAPE) can be used to compare different models and assess the accuracy of the models (Hyndman 2014, p.46; Hyndman & Koehler 2006).

The SEE measures the standard deviation around the regression line. A smaller SEE value will indicate that the regression line is closer to the data giving a better fit. The SEE is calculated using Equation 3-14 with *n* representing the sample size and *k* representing the number of independent variables (Berenson et al. 2012, p.534).

$$S_{YX} = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{n-k}} \qquad \textbf{Equation 3-14}$$

The CMD is also referred to as the goodness of fit where the proportion of variability of the dependent variable is described by independent variables (Vivaracho-Pascual et al. 2016). The CMD does not infer a relationship between the independent and dependent variables, it rather describes the amount of variation of the dependent variable that can be explained by the independent variables. A CMD value closer to one indicates a good fit of the model on the actual data (Devore & Farnum 2005, p.526; Field 2009, p.179; Hill & Lewicki 2006, p.395).

$$R^2 = \frac{\sum_{i=1}^{n}\left(\widehat{Y}_i - \bar{Y}_i\right)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \qquad \textbf{Equation 3-15}$$

A model can fit the data perfectly by adding more variables that makes the adjusted coefficient of multiple determination (ACD), ($R^2_{adjusted}$), an important parameter to report when comparing different regression models. The ACD takes both the sample size and number of independent variables into account and gives a more comparative regression value. A large difference between the CMD and ACD values indicates that adding more variables increased the amount of variance in the model and that some variables are insignificant. It is calculated using Equation 3-16 (Berenson et al. 2012, p.584; Devore & Farnum 2005, p.527).

$$R^2_{adjusted} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right] \qquad \textbf{Equation 3-16}$$

The MAE and RMSE are both parameters with dimensions that are used to assess the error between models predicting the same quality variable. The MAE and RMSE is calculated using Equation 3-17 and Equation 3-18 respectively (Azadi & Karimi-Jashni 2016; Hyndman 2014, p.46; Hyndman & Koehler 2006).

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}\left|Y_i - \widehat{Y}_i\right|}{n} \qquad \textbf{Equation 3-17}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\widehat{Y}_i - Y_i\right)^2}{n}} \qquad \textbf{Equation 3-18}$$

The RMSE increases with the increase in the magnitude of an error from an individual point, making it an unreliable parameter. Willmott & Matsuura (2005) observed that there exist no relationship between the RMSE and average error resulting in no clear interpretation of the RMSE value. Table 3-7 summarises the results from five different scenarios with four data points. The sum of the errors stays constant resulting in a constant MAE, while the square error increases due to large errors from single data points resulting in an increased RMSE. The MAE is therefore the preferred parameter to compare models predicting the same variable (Willmott & Matsuura 2005; Hyndman & Koehler 2006)

Table 3-7: Difference between MAE and RMSE adapted from Willmott & Matsuura (2005)

Variable	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Error for point 1	2	1	1	0	0
Error for point 2	2	1	1	0	0
Error for point 3	2	3	1	1	0
Error for point 4	2	3	5	7	8
$\sum \lvert e_i \rvert$	8	8	8	8	8
MAE	2	2	2	2	2
$\sum \lvert e_i \rvert^2$	16	20	28	50	64
RMSE	2	2.2	2.6	3.5	4

The MAPE is defined as a scale independent parameter that can be used to compare the error for models predicting different output variables (Azadi & Karimi-Jashni 2016). The MAPE values is still influence by the scale of the values and can only be used if the dataset does not contain any zero or near zero values, which will result in an inequality or extremely high MAPE values. A MAPE value close to zero indicates a higher model accuracy as the predicted values are closer to the actual (Hyndman 2014, p.47; Hyndman & Koehler 2006). Hyndman & Koehler (2006) also stated that the MAPE penalises positive errors heavier than negative errors. The author concluded that all the data points should be much greater than zero for the MAPE value to give a significant value. It is calculated using Equation 3-19.

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right| \times 100 \qquad \textbf{Equation 3-19}$$

Chapter 4. Results and discussion

The screening phase (phase one) results were used to determine the significance of the chosen independent variables to model the dependent variables. The accuracy and performance of the multiple linear regression (MLR) technique was evaluated in section 4.1.2 where only main effects models (MLR-M) were develop. The Spearman's rho matrix described the correlations between the operating and output variables, and is discussed for each phase in sections 4.1.1 and 4.2.1.

Main effects and two-way interaction effects (MLR-I) multiple linear regression models were developed in the modelling phase (phase two), and is discussed in section 4.2.2. The main effects and interaction effects models were compared to determine the models with the best performance. The effect of added independent variables was also investigated.

Artificial neural networks (ANNs) were investigated as a non-linear comparison technique. The ANNs are only a qualitative study and were developed using rules-of-thumb. The interaction effects MLR models were compared to ANN models developed using all the independent variables as inputs (ANN-I) in section 4.2.3.

The use of the main effects MLR technique to eliminate unnecessary independent variables and reduce the complexity of the ANN models were evaluated in Appendix C. ANN models (ANN-M) were developed using the same independent variables as those obtained from the main effects MLR models developed in section 4.2.2. These main effects ANN models were compared with the ANN models, using all the independent variables as inputs, in Appendix C4.2.3.

4.1 Results from phase one

The Spearman's rho matrix was used to determine the relationships between the operating and output variables. The results from the Spearman's rho matrix served as a preliminary step to determine the possible independent variables that can be used to model a dependent variable. The main effects MLR model's performance and accuracy were statistically evaluated.

4.1.1 Spearman's rho matrix

The Spearman's rho matrix is divided into three smaller matrixes. The full Spearman's rho results and the p-values for each correlation are summarised in Table D-1 and Table D-2 in Appendix D. The first matrix is used to assess the independency of the operating variables.

The correlation values for the first Spearman's rho matrix are summarised in Table 4-1, where the asterisk denotes a 95% statistical certainty and the double asterisks a 99% statistical certainty.

Table 4-1: Frist Spearman's rho matrix for phase 1

	FAF	FAT	SLF	SLT	SLC	SPS	SSL
FAF	1	-.31	.39*	.23	.01	.40*	.13
FAT	-.31	1	-.33	-.02	-.13	.08	.34*
SLF	.39*	-.33	1	.32	-.28	.42*	-.23
SLT	.23	.02	.32	1	.15	.08	-.09
SLC	.01	-.13	-.28	.15	1	-.14	.15
SPS	.40*	.08	.42*	.08	-.14	1	.51**
SSL	.13	.34*	-.23	-.09	.15	.51**	1

Note: *p<.05, **p<.01

The correlations between the fluidising air flow rate (FAF) with the spray liquid flow rate (SLF) and the seed mean particle size (SPS) variables are the largest correlations for the FAF with values of .39, p (two-tailed) < .05, and .40, p (two-tailed) < .05, respectively. There exist no physical connection between these variables and these values are not high enough to indicate a significant correlation.

The largest negative correlation in the matrix is obtained between the fluidising air temperature (FAT) and SLF with a value of -.33, with a confidence lower than 95%. Both variables are operated independently from one another and the value is too low to have any significance. The correlation between the FAT and seed particle slope (SSL) has a value of .34, p (two-tailed) < .05. The value is also not high enough to imply a relation between these variables.

The spray liquid temperature (SLT) and spray liquid concentration (SLC) variables have low correlations with regards to the other operating variables. The largest correlations for these variables are obtained between the SLT and SLF with a value of .32, with a confidence lower than 95%, and the SLC and SLF with a value of -.28, with a confidence lower than 95%. Both values are small and have p-values larger than .05, making these results statistically insignificant.

The SPS variable has the largest positive correlation in the matrix with the SSL variable with a value of .51, p (two-tailed) < .01. The relation between the SPS and SSL is due to the adjustments made on the crusher in order to achieve different SPS values. This also affected the particle size distribution (PSD) and in turns the SSL. The correlations between the SSL and the FAT and SPS variables are concluded to be insignificant.

None of the correlations were high enough to indicate a significant correlation. The only fairly significant correlation obtained was between the SPS and SSL, both being particle size properties, which is attributed to the interdependency observed in the crusher producing the seed particles. In conclusion, the results from the first Spearman's rho matrix support the assumption of having independent operating variables.

The second Spearman's rho matrix describes the relationships between the independent and dependent variables. These correlations give an indication which independent variables will be significant during the modelling phase. The values for the second Spearman's rho matrix are presented in Table 4-2.

Table 4-2: Second Spearman's rho matrix for phase 1

	FAF	FAT	SLF	SLT	SLC	SPS	SSL
RUR	.40*	-.39*	.96**	.31	-.26	.35*	-.26
RER	-.45**	.48**	-.76**	-.31	-.06	-.52**	.15
EFF	.40*	-.37*	.51**	.69**	.35*	.08	-.28
PPS	.13	.06	.29	.06	.09	.75**	.33
GPS	.32	.02	.66**	.06	-.23	.72**	.20
POR	.09	-.25	.17	-.14	-.02	-.13	-.18
PCR	.08	.49**	-.29	-.13	-.37*	.16	.41*
PSL	-.14	.22	-.77**	-.04	.21	-.30	.30
GSL	.27	.14	-.46**	-.18	.01	.02	.34*

Note: *p<.05, **p<.01

The FAF has a positive correlation with the production rate (RUR) and efficiency (EFF) variables, both with a value of .40, p (two-tailed) < .05. An increase in the FAF results in an increase in the drying rate of the spray liquid according to Fries et al. (2014). A shorter drying period increases the EFF and increases the RUR. There is also a negative correlation between the FAF and recycle ratio (RER) with a value of -.45, p (two-tailed) < .01. This correlation is explained by the increase in the RUR resulting in a lower number of product that is recycled.

The FAF also has a positive correlation with the granulator mean particle size (GPS) and granulator particle slope (GSL) variables, with values of .32 and .27. Fries et al. (2014)

concluded than an increase in the FAF will increase the attrition rate, resulting in the formation of a larger number of small granules. The positive Spearman's rho correlation for the GPS variable contradicts this finding from literature. The correlation between the FAF and GSL is more in line with literature observations, where a higher slope indicates an increase in the number of granules inside a specific size range. Both rho values for the GPS and GSL variables have a confidence lower than 95% due to the absence of the asterisks next to the values, resulting in a low confidence for the correlation values.

The FAT has low negative correlations with the RUR and EFF variables, with values of -.39, p (two-tailed) < .05, and -.37, p (two-tailed) < .05, respectively. These correlations are almost the opposite as those obtained between FAF variable and the RUR and EFF. There should be no correlation between the FAT and RUR according to Srinivasakannan & Balasubramaniam (2003). The higher FATs make operation at higher liquid flow rates possible but do not influence the growth rate of the granules (Srinivasakannan & Balasubramaniam 2003). Ziyani & Fatah (2014) found that an increase in the FAT resulted in the formation of smaller granules. The negative correlation between the FAT and RUR could be explained by this observation, were more undersize particles are produced as a result of the increased FAT. Becher & Schlünder (1998) concluded that increasing the FAT will increase the evaporation rate, resulting in an increased spray efficiency and decrease in granule size. The correlation obtained between the FAT and EFF variables are contradicting these observations from literature. The correlation value is dependent on the size and levels of the dataset. This correlations might change when a larger dataset with more levels are used.

The RER and product circularity (PCR) have positive correlations with the FAT variable with rho values of .48, p (two-tailed) < .01, and .49, p (two-tailed) < .01. Biswal (2011) observed that an increase in the FAT favours the layering mechanism and produces smaller granules. The layering mechanism produces smaller spherical granules, resulting in higher circular granules and larger recycle rates due to the smaller granules.

The SLF has a very strong correlation with the RUR variable with a value of .96, p (two-tailed) < .01. Fries et al. (2014) and Srinivasakannan & Balasubramaniam (2003) concluded that an increase in the SLF increases the droplet size, resulting in the production of larger granules and increased granule growth. The SLF also has positive correlations with the EFF and GPS variables, having correlation values of .51, p (two-tailed) < .01, and 66, p (two-tailed) < .01, respectively. The positive RUR and GPS correlations agree with the observations from literature, where larger granules are produced at a larger production rate. The increase in the

spray rate results in a higher amount of spray liquid that forms part of the final granules, explaining the positive EFF correlation obtained.

The RER and PSL variables have strong negative correlations with the SLF with values of -76, p (two-tailed) < .01, -.77, p (two-tailed) < .01. The increase in flow rate results in an increase in production, therefore resulting in a decrease in the RER. There is a low negative correlation between the SLF and the GSL variables with a value of -.46, p (two-tailed) < .01. An increase in the SLF increase the granule growth rate, resulting in an even particle size distribution. This explained the negative correlation between the SLF and GSL.

The SLT has moderately low correlations with the other output variables, apart from a strong positive correlation with the EFF variable with a value of 69, p (two-tailed) < .01. Sahoo (2012) observed that an increase in the SLT decreases the viscosity and droplet size. The drying period is lower for smaller droplets at higher temperatures, resulting in an increase in the spray efficiency. This correlation is therefore supported from the observation from literature.

An increase in the SLC favours the agglomeration mechanism resulting in an increase in granule size and production rate according to Sahoo (2012) and Srinivasakannan & Balasubramaniam (2003). The SLC has negative correlations with the RUR, GPS and PCR variables. The negative correlations between the SLC variable and the RUR and GPS variables are contradicting to these literature observations. Both correlations values between the RUR and GPS variables with the SLC are low with p-values higher than .05, making these correlations insignificant. Only the negative correlation between the SLC and PCR variable is significant, with a correlation value of -.37, p (two-tailed) < .05. Agglomerates have an asymmetrical shape resulting in a reduced product circularity. This explains the negative correlation obtained between the SLC and PCR. A low positive correlation value of .35, p (two-tailed) < .05, is obtained between the SLC and EFF variables. An increase in the spray concentration increases the amount of ammonium nitrate in the spray liquid, thus resulting in an increase in the EFF.

The SPS has strong positive correlations with the product mean particle size (PPS) and GPS variables, with values of .75, p (two-tailed) < .01, and 72, p (two-tailed) < .01, respectively. The seed particles form the basis upon which the product and granulator granules are grown. Increasing the seed particle size will result in larger product and granulator granules. An increase in the SPS also has a positive correlation with the RUR, while having a negative correlation with the RER variable. The correlation values between the SPS and the RUR and RER variables are .35, p (two-tailed) < .05, and -.52, p (two-tailed) < .01, respectively. Granule growth is dependent on different operating variables and conditions. Biswal (2011) and

Srinivasakannan & Balasubramaniam (2003) concluded that a increase in the SPS favours the layering mechanism resulting in a lower growth rate. The SPS positive correlations with the RUR and negative correlation with the RER contradicts this observation from literature.

The SSL has positive correlations with the PCR and GSL variables with value of .41, p (two-tailed) < .05, and .34, p (two-tailed) <.05. It also exhibits a positive correlation with the product particle slope (PSL) variable with a value of .30, but with a confidence lower than 95%. Higher seed particle slopes increases the particle size fraction that experiences the highest granule growth, resulting in an increase in the granulator and product slope values.

The SLT and SSL have low correlations with most of the output variables. These values might not be included in most models during the modelling phase. The product porosity (POR), PCR and GSL output variables show very low correlations with most operating variables. This lack of correlations could make it difficult to develop statistical significant models. Most of the results obtained in this matrix agree with the findings in literature. The correlation results are dependent on the dataset. A larger dataset with more levels will improve the significance of the results, leading to more correlations agreeing with literature observations.

The third matrix is helpful to determine the effects that can be expected if one quality variable is optimised. The results are given in Table 4-3.

Table 4-3: Third Spearman's rho matrix for phase 1

	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
RUR	1	-.69**	.60**	.18	-.58**	.12	-.30	-.75**	-.46**
RER	-.69**	1	-.49**	-.52**	-.52**	-.18	.43*	.61**	.39*
EFF	.60**	-.49**	1	.07	.14	-.05	-.50**	-.30	-.38*
PPS	.18	-.52**	.07	1	.70**	-.21	-.05	-.21	-.10
GPS	.58**	-.52**	.14	-.70**	1	-.07	.06	-.42*	-.09
POR	.12	-.18	.05	-.21	-.07	1	-.12	-.17	-.15
PCR	-.30	.43*	-.50**	-.05	.06	-.12	1	.32	.47**
PSL	-.75**	.61**	-.30	-.21	-.42*	-.17	.32	1	.49**
GSL	-.46**	.39*	-.38*	-.10	-.09	-.15	.47**	.49**	1

Note: *p<.05, **p<.01

The RUR and PSL have the highest negative correlation in the table, with a value of -.75, p (two-tailed) < .01. The results from the second matrix showed opposite correlation values obtained between each operating variable and the RUR and PSL variables. This already indicated that there is a strong negative correlation between the two dependent variables. There is also a strong positive correlation between the RUR and GPS variables with a value of .58, p (two-tailed) < .01. The RUR also has large negative correlations with the RER and

GSL, with correlation values of -.69, p (two-tailed) < .01, and -.46, p (two-tailed) < .01. An increase in the SLF, increases the granule size, GPS, and production rate, RUR (Srinivasakannan & Balasubramaniam 2003). An increase in the RUR is accompanied by a decrease in the RER in order to keep the mass balance of the process in check. An increase in granule growth results in an even particle size distribution, therefore a decrease in the particle size distribution slope. This explains the negative between the RUR, PSL and GSL variables with the RUR and the positive correlation between the GPS and RUR.

The RER variable has high negative correlations with the EFF, PPS and GPS variables, with values of -.49, p (two-tailed) < .01, -.52, p (two-tailed) < .01, and -.52, p (two-tailed) < .01, respectively. The recycle stream consists of granulator fines and crushed oversize granules. The production of very small granules will result in an increase in the RER, a significant increase in the granule size will also result in an increase in the RER. The granulator product is sieved to remove the under and oversize particles. The on-size particles form the final product stream and the production rate is represented by the RUR. This means that the RUR is also negatively influenced by the production of very fine or very large granules. The PPS and GPS variables have non-monotonic relationships with the RUR and RER variables. The Spearman's rho correlation assesses monotonic behaviour, making the rho values for these correlations irrelevant. An increase in the spray efficiency, EFF, will result in an increase in granule growth, which increases production and reduces the RER. There are also positive correlations between the RER and the PCR, PSL and GSL variables, with rho values of .43, p (two-tailed) < .05, .61, p (two-tailed) < .01, and .39, p (two-tailed) < .05, respectively. Higher particle size distribution slopes indicates a large fraction of particles within a specific size range. An increase in the RER will be obtained, if the particle size range is larger than the on-size particle size range.

The EFF and RUR have a positive correlation with a value of .60, p (two-tailed) < .01, where an increase in spray efficiency results in an increase in production. An increased EFF correlates negatively with the PCR with a value of -.50, p (two-tailed) < .01. An increased EFF is mainly due to an increased in SLF and SLC which favours the agglomeration mechanism, resulting in the production of irregular shaped granules that explains the negative correlation (Srinivasakannan & Balasubramaniam 2003; Wong et al. 2013).

The PPS and GPS have the strongest positive correlation of .70, p (two-tailed) < .01. This result is due to the granulator product being sieved to form the final product. The size distribution properties of the granulator and product samples are closely related to one another.

The POR variable exhibits low correlations with the other dependent variables. The highest POR correlation obtained is between the POR and the PPS variable with a rho value of -.21, with a confidence lower than 95%. Agglomerates have larger particle sizes with higher porosities compared to layered granules. This explanation supports the Spearman rho correlation between the POR and PPS value obtained.

4.1.2 Multiple linear regression

For phase one, main effects multiple linear regression (MLR) were investigated to model the output variables. The coefficients for the final main effects models are summarised in Table D-5 in Appendix D. The statistical parameter results for the main effects models are summarised in Table 4-4.

Table 4-4: Main effects multiple linear regression model results for phase 1

	RUR	RER	EFF	PPS	GPS	PCR	PSL	GSL
Model CMD	0.97	0.86	0.81	0.51	0.62	0.61	0.73	0.44
SEE	0.21	0.40	1.43	0.08	0.06	0.39	0.44	0.37
MAPE	3.1%	6.5%	1.4%	3.0%	3.1%	0.3%	5.7%	7.0%
Normal distribution	Yes	Yes	No	Yes	Yes	Yes	Yes	No

The production rate (RUR), recycle ratio (RER) and efficiency (EFF) showed a fair coefficient of multiple determination (CMD) of at least 80% with low standard error of estimate (SEE) values, resulting in a good fit of the actual data. The mean absolute percentage error (MAPE) values are moderately low for the RUR and EFF models indicating a fair accuracy. Figure 4-1 shows that the RUR model follows the trend of the data fairly well with the exception of a few points.

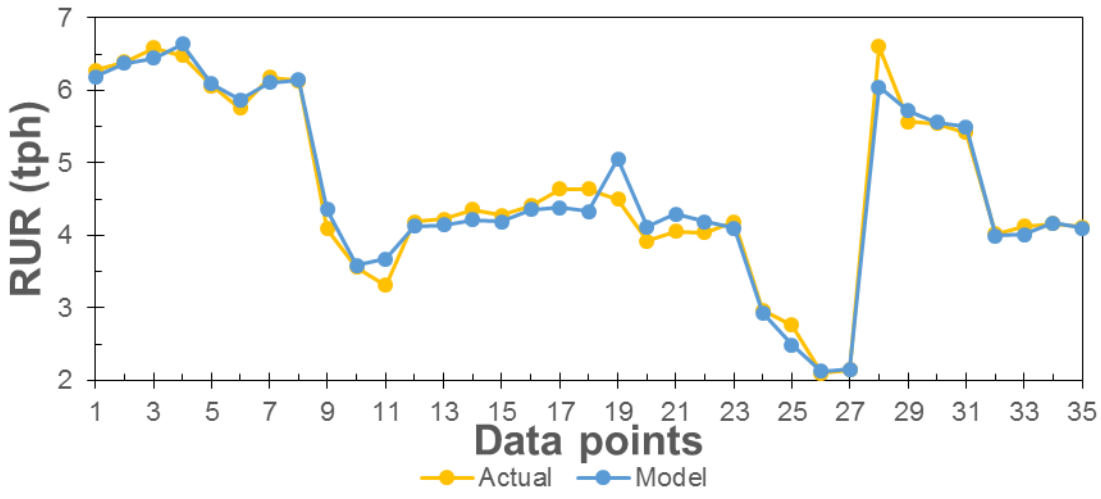


Figure 4-1: Phase 1 MLR main effects RUR model performance

The actual data points for the RER variable does not vary in a continuous way, resulting in a more discrete-like dataset as shown in Figure 4-2. The model predictions are distributed around the actual variables, especially in the 1 to 8 and 9 to 19 data point range. The moderate MAPE value of 6.5% indicates a low accuracy for the RER model. The discrete-like data resulted in a fair prediction fit by the model but with a low accuracy.

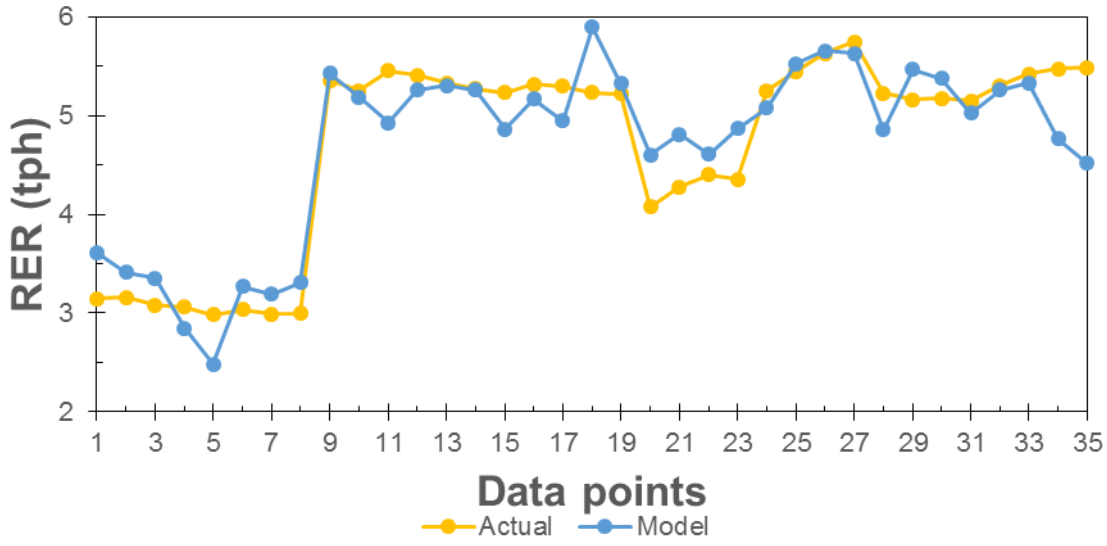


Figure 4-2: Phase 1 MLR main effects RER model performance

The EFF model fits the data reasonably well as shown in Figure 4-3, but struggles to fit all the points with a high accuracy resulting in the moderate SEE value of 1.43. The EFF model follows the trend of the data well. All three models show sufficient performances based on their high regression values and moderate accuracies based on their low error values.

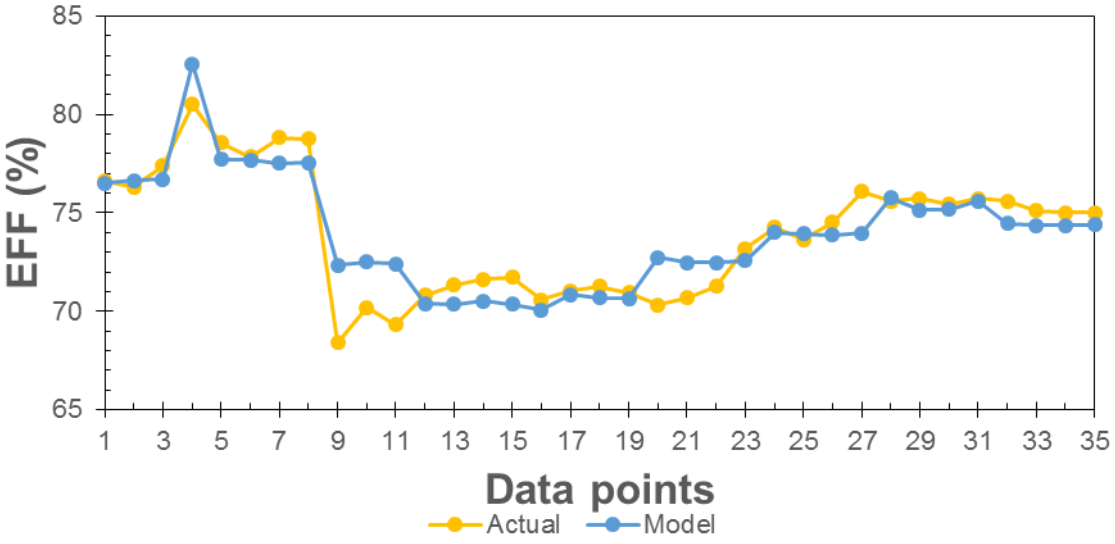


Figure 4-3: Phase 1 MLR main effects EFF model performance

The product mean particle size (PPS) and granulator mean particle size (GPS) models have low regression values along with low SEE values. Both models show visual adequate performance following the actual data with low deviations resulting in the low SEE values. The moderate MAPE values indicate large errors on some data points, e.g. data points 6, 7 and 17 for the PPS model and data points 2, 5, 13 and 14 for the GPS model as perceived in Figure 4-4 and Figure 4-5. The models follow the trend of the data moderately well, but with moderate accuracies resulting in inadequate models.

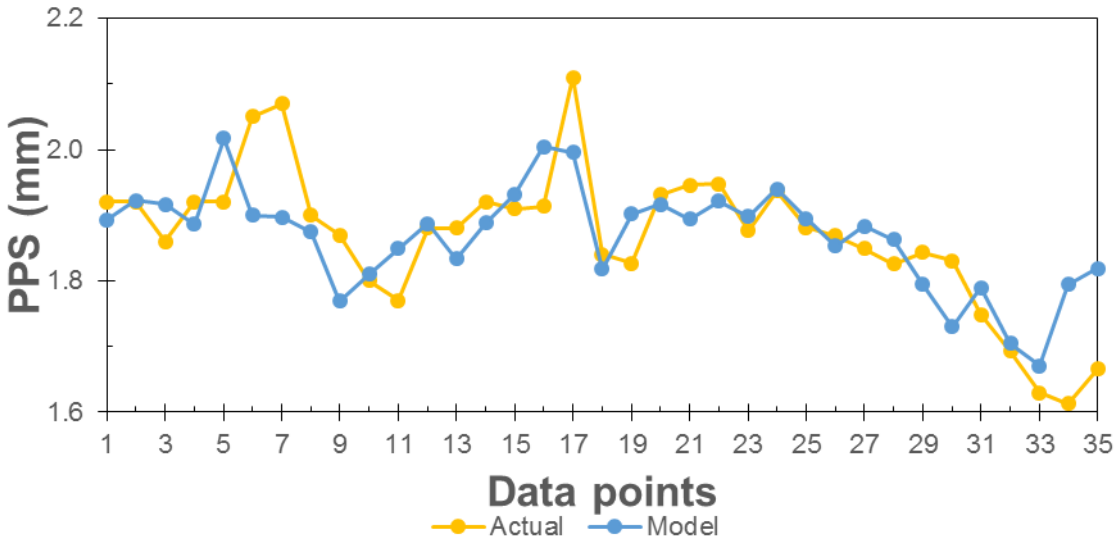


Figure 4-4: Phase 1 MLR main effects PPS model performance

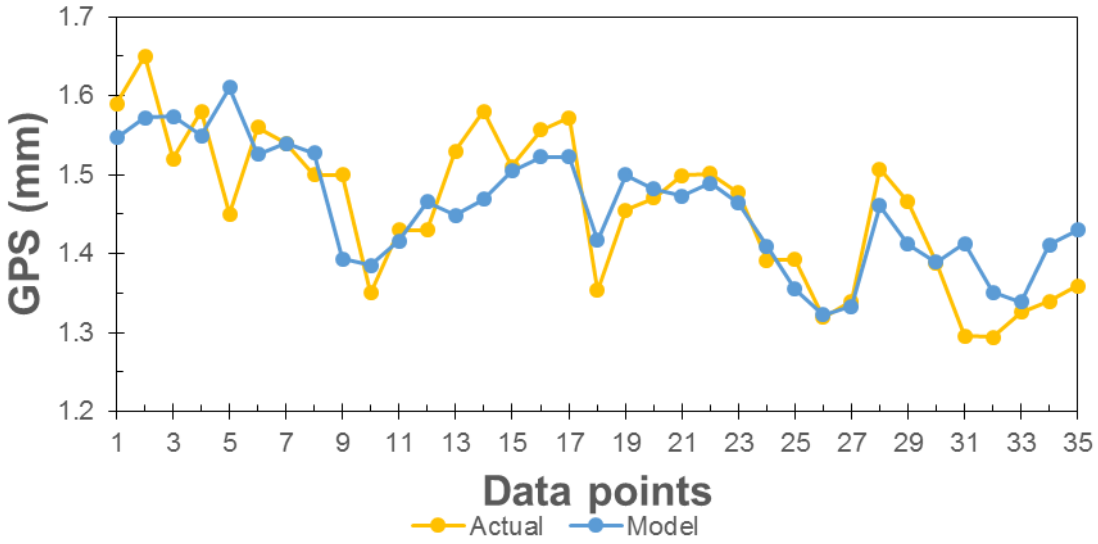


Figure 4-5: Phase 1 MLR main effects GPS model performance

The product circularity (PCR) has a low CMD of 0.61, with reasonably low MAPE and SEE values of 0.3% and 0.39 respectively. The predicted values seem to follow the general tendency; however, there exist a large offset between the actual and predicted values as observed in data points 9 to 12 in Figure 4-6. The model follows the general tendency of the data but with low accuracies at data points 2, 9 to 12 and 31.

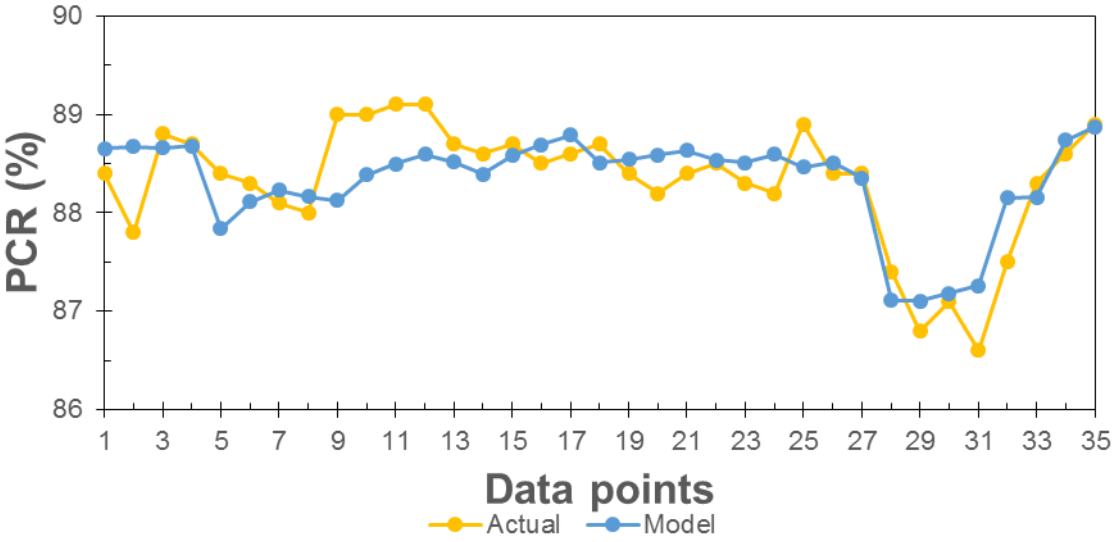


Figure 4-6: Phase 1 MLR main effects PCR model performance

The product particle slope (PSL) model has a high CMD value of 73% with a moderate MAPE value of 5.7%. The model follows the general trend of the data well but with a low prediction accuracy as shown in Figure 4-7. Large error values are obtained at data points 7, 16 and 17, resulting in the moderate MAPE value obtained. The model cannot be used for prediction due to inaccurate predictions and relative high errors.

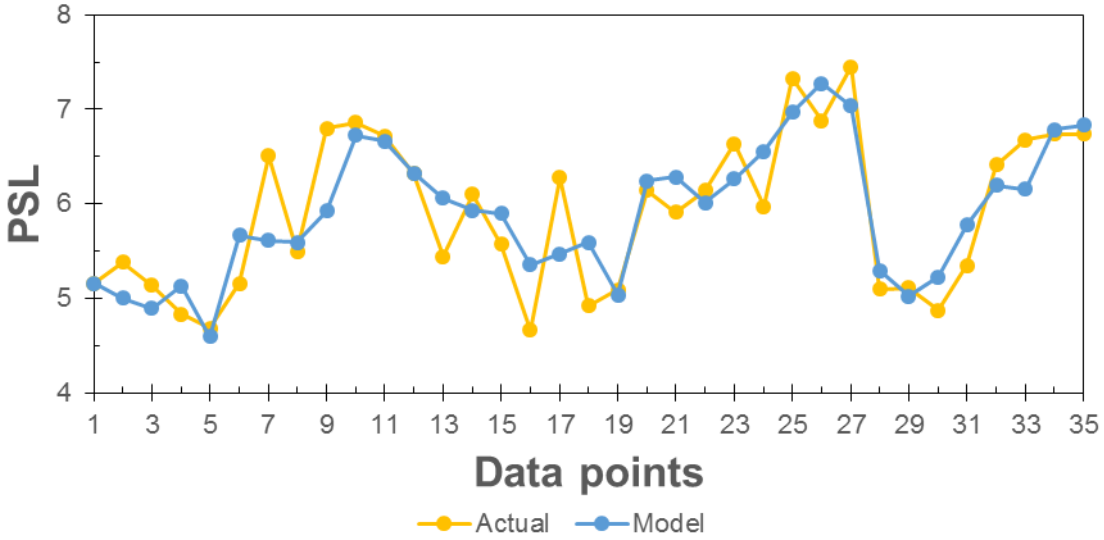


Figure 4-7: Phase 1 MLR main effects PSL model performance

The granulator particle slope (GSL) has the lowest CMD value and the highest MAPE value in the table. The model fit only about all the data points in the second half of the dataset, as observed in Figure 4-8. The errors from the data points in the first half, especially data points 1, 5 and 7 to 10, contribute to the high MAPE value obtained. The model is concluded inaccurate for prediction due to the lack of an overall predictability.

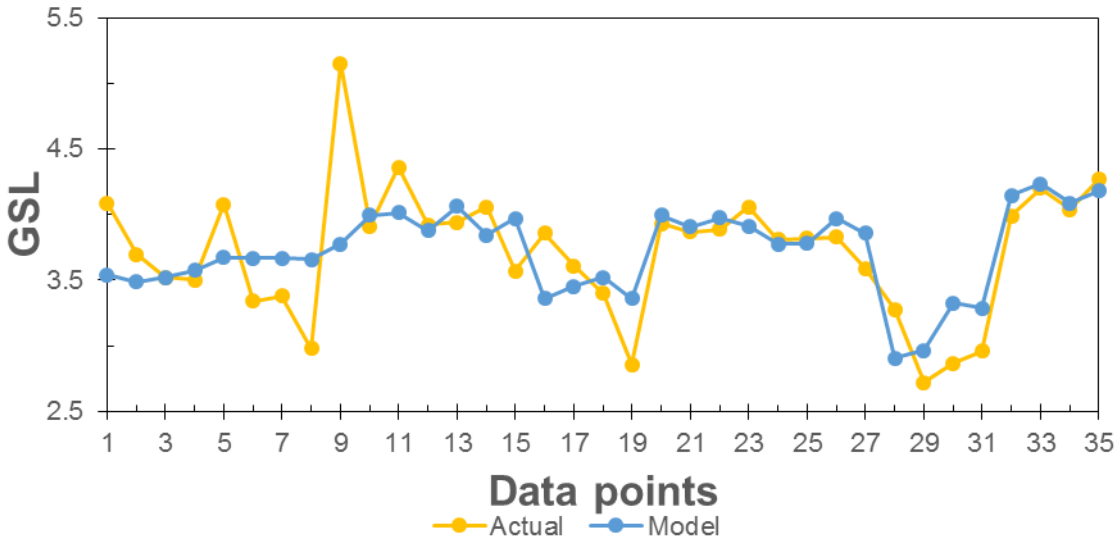


Figure 4-8: Phase 1 MLR main effects GSL model performance

A statistical significant main effects MLR model could not be developed for the product porosity (POR). The Spearman results indicated low correlations between the operating variables and the POR confirming the lack of a statistical significant main effects model. A two-way interaction effects MLR model was developed and evaluated. The statistical significant coefficients for the POR model are summarised in Table D-5 in Appendix D. The results from the model are summarised in Table 4-5.

Table 4-5: Interaction effects multiple linear regression model results for phase 1

	POR
Independent variables	7
Model CMD	0.47
ACD	0.31
SEE	3.70
MAPE	10.6%
Normal distribution	Yes

The interaction effects model has a low fit with a CMD value of 47%. The model visibly seems to fit the data reasonably well, but with a low accuracy as shown in Figure 4-9. A large difference between the CMD and the adjusted coefficient of multiple determination (ACD) indicates that including more variables, added more variance into the model. The model is greatly influenced by the addition of variables, indicating that some of these variables do not contribute significantly to the model. Removing these variables will result in a lower CMD value and a lower model performance. The model also has a high MAPE value of 10% and a high SEE value of 3.70. The model struggles to follow the extreme data points, as found with data points 20, 24 and 33, resulting in inaccuracies and large errors.

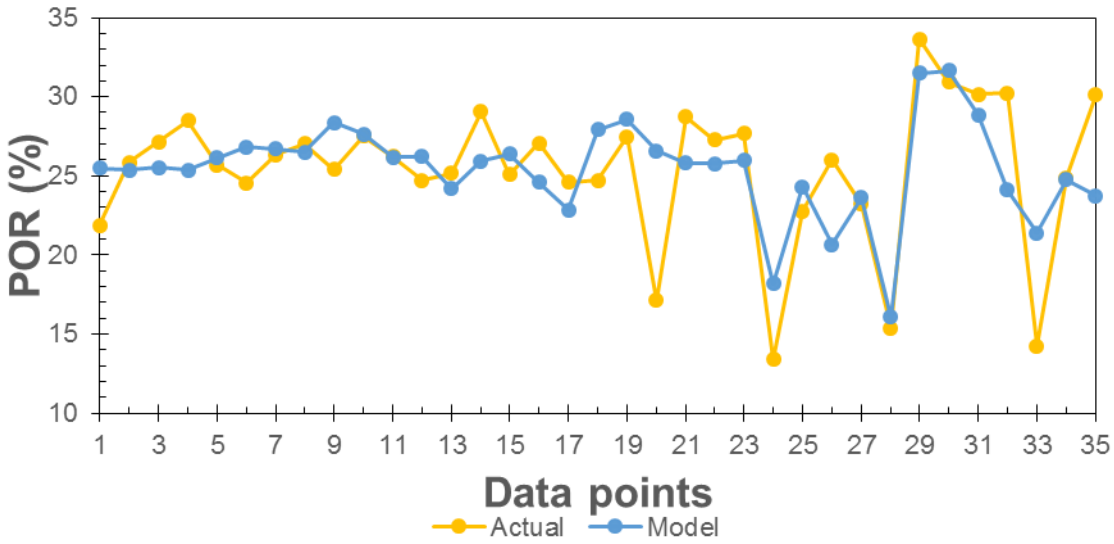


Figure 4-9: Phase 1 MLR interaction effects POR model performance

The results from the Shapiro-Wilk test concluded that all the residual values of the models, apart from the EFF and GSL models, were normally distributed. The MLR modelling technique assumes that the residuals fit a normal distribution; however, this assumption is not accepted with the EFF and GSL models.

The Spearman's rho results were useful in determining the expected relations and the possible independent variables for model development. Even though most of the models exhibit normal distribution, the results from these models indicate that more data are required to develop models with a higher accuracy. Evaluating the variables at more levels will also increase the significance of the Spearman's correlation and the developed models.

Most of the models showed a good fit with sufficient accuracies. These models are developed with a small dataset resulting in a low confidence in the usability of these models. A larger dataset with more levels will increase the accuracy of the usability of the models. Including

two-way interaction effects for model development can increased the accuracies and performances of the models. Interaction effects are expected due to the multiple influences from the operating variables on the growth mechanism. The growth mechanism plays an important role in the product formation and product quality.

4.2 Results from phase two

The second phase was evaluated on a larger dataset with more levels as discussed in section 3.2.2. The Spearman matrix constructed in phase one concluded the independency of the operating variables, resulting in including all the chosen operating variables in the modelling phase. The Spearman's rho values from the second phase were also useful in determining the correlations between the variables. Most of the results obtained compared with results obtained from the screening phase and with the observations from literature.

Main effects (MLR-M) multiple linear regression (MLR) and interaction effects MLR models (MLR-I) were developed as described in section 3.2.2 and compared in section 4.2.2 to assess the inclusion of two-way interaction variables on the accuracy and performance of the MLR models.

Artificial neural network (ANN) models were developed to qualitatively compare with the MLR modelling technique. The difference between the MLR and non-linear ANN modelling techniques were evaluated in section 4.2.3 by comparing the interaction effects MLR models with the ANN models that were developed using all the independent variables (ANN-I).

The use of the main effects MLR models to develop smaller ANN models were developed in Appendix C. The independent variables from the final main effects MLR models from section 4.2.2 were used to develop smaller ANN models (ANN-M). These models were compared with ANN models that included all the independent variables as developed in section 4.2.3.

4.2.1 Spearman's rho matrix

The full Spearman's rho results and the p-values for each correlation are summarised in Table D-3 and Table D-4 in Appendix D. The independency of the operating variables is once again evaluated by the first Spearman's rho matrix summarised in Table 4-6.

Table 4-6: Frist Spearman's rho matrix for phase 2

	FAF	FAT	SLF	SLT	SLC	SPS	SSL
FAF	1	-.07	-.10	.04	.05	-.02	-.18
FAT	-.07	1	-.16	-.04	.09	-.26*	.51**
SLF	-.10	-.16	1	-.05	-.21	.44**	-.22
SLT	.04	-.04	-.05	1	.75**	-.36**	-.02
SLC	.05	.09	-.21	.75**	1	-.44**	-.04
SPS	-.02	-.26*	.44**	-.36**	-.44**	1	-.09
SSL	-.18	.51**	-.22	-.02	-.04	-.09	1

Note: *p<.05, **p<.01

The fluidising air flow rate (FAF) has low correlation values with the other operating variables. The Spearman's rho results from the first phase indicated correlations between the FAF and the spray liquid flow rate (SLF) and seed mean particle size (SPS) variables. The larger dataset and increased experimental levels lowered these correlation values and significance by increasing the variance.

The fluidising air temperature (FAT) has a positive correlation of .51, p (two-tailed) <.01, with the seed particle slope (SSL). This correlation was also observed from the Spearman results from the first phase, however, the correlation between the FAT and SLF is not present in the second phase. There exists no physical connection between the FAT and SSL variables. The only possible explanation for the correlation could be due to the influence that the FAT has on the granulator product. This influence will be present in the recycled granules and could therefore influence the SSL.

A positive correlation is also observed between the SLF and SPS with a value of .44, p (two-tailed) < .01. This correlation along with a correlation between the SLF and spray liquid temperature (SLT) were observed in the screening phase Spearman's rho results. This correlation can also be explained by the influence that the SLF has on the granulator product which is recycled back, forming the seed material. It is important to note that the SLF could influence the SPS due to the seed material consisting of recycled material, but the SPS cannot influence the SLF.

The largest positive correlation obtained is between SLT and spray liquid concentration (SLC) with a value of .75, p (two-tailed) < .01, indicating a strong positive correlation between these variables. The SLT value was calculated for each concentration set in order to prevent nozzle blocking as explained in section 3.2.2. The SLT values had to increase as the SLC increased to ensure plant operation, resulting in the correlation obtained.

There is also a low negative correlation between the SLT and SPS with a value of -.36, p (two-tailed) < .01. The largest negative correlation exists between the SLC and SPS with a value of -.44, p (two-tailed) < .01. The fluidising and spraying variables are completely independent from any other operating variable, but the seed particles consist of recycled material. The recycled granules are influenced by the other operating variables, resulting in a non-direct influence from the fluidising and spraying variables on the seed particle variables.

Due to the low correlations, apart from the SLC and SLT correlation, it can be concluded that the operating variables are independent from one another. The Spearman's rho values from the second phase have higher p-values compared to the first phase, indicating that the larger dataset and more experimental levels improved the significance of the correlation values.

The correlations between the operating and output variables are summarised in Table 4-7.

Table 4-7: Second Spearman's rho matrix for phase 2

	FAF	FAT	SLF	SLT	SLC	SPS	SSL
RUR	-.06	-.14	.74**	.17	.12	.30*	-.05
RER	-.00	.21	-.26*	-.55**	-.41**	.13	.04
EFF	-.05	-.09	-.04	.71**	.72**	-.26	-.08
PPS	-.16	-.27*	.10	.42**	.28*	-.12	-.37**
GPS	-.08	-.09	.16	.12	-.09	.26*	-.13
POR	-.03	.22	.14	-.23	.03	-.15	.17
PCR	-.10	.01	.01	-.63**	-.49**	.17	-.09
PSL	-.28*	-.04	-.39**	-.15	-.21	-.14	.17
GSL	.01	.55**	-.39**	.35**	.20	-.36**	.47**

Note: *p<.05, **p<.01

The FAF has low correlations with the other output variables. The correlation between the PSL variable is the largest with a rho value of -.28, p (two-tailed) < .05. Smith & Nienow (1983) concluded that an increase in the FAF would increase the attrition rate decreasing the particle size and favouring of the layering mechanism. The increase in attrition rate will result in a uniform particle size distribution, therefore resulting in a lower particle size slope. However, the correlation value is too small to make any significant conclusions.

The results from the first phase showed correlations between the FAF and the production rate (RUR), recycle ratio (RER) and efficiency (EFF) variables with correlation values in the order of .40. The increase in dataset and experimental levels resulted in the absence of these correlations in the second phase, indicating a low influence from the FAF variable.

The FAT has a positive correlation with the granulator particle slope (GSL) variable with a value of .55, p (two-tailed) < .01. An increase in the FAT increases the evaporation rate

resulting in the production of smaller granules (Fries et al. 2014). The increase of small granules results in a narrow particle size distribution with a high particle slope value, explaining the correlation obtained. The Spearman results from the first phase indicated correlations between the RUR, RER, EFF and product circularity (PCR) variables. None of these correlations is found to be significant in the second phase.

The SLF and RUR variables have the largest positive correlation in the table with a value of .74, p (two-tailed) < .01. This correlation was also present in the results from phase one. An increase in the SLF increases the droplet size and growth rate resulting in an increase in the RUR (Fries et al. 2014; Srinivasakannan & Balasubramaniam 2003). The increased flow rate favours agglomeration and the formation of larger granules. This supports the correlations between the SLF and the product particle slope (PSL) and GSL variables with both having the same value of .39, p (two-tailed) < .01. The rho value between the SLF and PSL decreased compared to the results from phase one, reducing the strength of the correlation.

The strong negative correlation between the SLF and RER variables from phase one is not present in phase two. The strength of the correlation is lower with a value of -.26, p (two-tailed) < .05. The large positive correlations between the SLF and the EFF and granulator mean particle size (GPS) variables from phase one are also absent in the second phase. The SLF variable became less significant in the second phase, with only a few strong correlations.

The SLT has strong correlations with the EFF and PCR output variables, with correlation values of .71, p (two-tailed) < .01, and -.63, p (two-tailed) < .01, respectively. The strong EFF correlation was observed in the phase one Spearman's rho result, but the strength of the correlation for the PCR increased with the larger dataset. Sahoo (2012) found that an increase in the temperature decreases the spray liquid viscosity and droplet size. The higher temperatures will also decrease the drying period resulting in faster growth and an increased EFF. The decreased viscosity will results in the formation of weak liquid and solid bridges, resulting in the production of friable granules with a low circularity, explaining the negative correlation between the SLT and the PCR.

An increase in the EFF will result in an increase in production and a decrease in the RER, explaining the correlation between the SLT and RER with a value of -.55, p (two-tailed) < .01. The rho value between the SLT and RER increased significantly compared to the phase one Spearman's rho results. The positive correlations between the SLT and the product mean particle size (PPS) and GSL variables are not found in the second phase. Stronger correlations between the SLT and the output variables are observed in the second phase

where only a strong correlation between the SLF and EFF was obtained in the first phase. The SLT variable became more significant in the second phase.

The SLC also has a strong correlation with the EFF variable with a value of .72, p (two-tailed) < .01. A spray liquid with a high concentration favours agglomeration and an increase in the growth rate, increasing the RUR as concluded by Sahoo (2012). Less water is present in the spray liquid, resulting in an increase the efficiency of the granulator. The increased RUR results in a decrease in the RER, confirming the negative correlation of -.41, p (two-tailed) < .01, between the SLC and RER. The opposite correlation was obtained between the SLC and RUR in phase one. The correlation between the SLC and RUR from phase two agrees with the observations from literature. This indicates that the increase dataset and experimental levels improved the significance of the Spearman's rho results. Both rho values for the RER and EFF correlations increased significantly compared to the first phase. The negative correlation between the SLC and PCR also increase to a value of -.49, p (two-tailed) < .01. An increase in the SLC favours the agglomeration mechanism resulting in the production of irregular shapes granules, decreasing the PCR.

The SPS has low positive correlations with the RUR and GPS variables with values of .30, p (two-tailed) < .05, and .26, p (two-tailed) < .05. The RUR correlation contradicts the conclusions from literate where an increase in the SPS favours the layering mechanism and decrease the growth rate (Biswal 2011; Srinivasakannan & Balasubramaniam 2003). The increased RUR could be because of a larger fraction of on-size particles due to the production of smaller particles. The same correlation between the SPS and RUR was obtained in the first phase with a higher value of .35, p (two-tailed) < .05.

The correlation values for the PPS and GPS variables significantly decreased compared to the Spearman's rho results from the screening phase. The PPS has a negative correlation, which contradicts the observations from literature and from the first phase. The correlation between the SPS and PPS has rho value of -.12, with a p-value larger than .05, making the correlation insignificant. An increase in the SPS correlates with a decrease in the PSL and GSL. The rho values between the SPS and the PSL and GSL variables are -.14, with a p-value larger than .05, and -.36, p (two-tailed) < .01. An increase in the SPS results in a wider particle size distribution resulting in a lower slope value.

The SSL has a positive correlation with the PSL and GSL variables. The only significant correlation is with the GSL variable with a value of .47, p (two-tailed) < .01. A positive correlation between the slope variables can be expected since an increase in the SSL increases the particle size fraction that experiences the most granule growth. These

correlations were also observed in the first phase, but with a larger rho value for the PSL variable.

The SSL also correlates negatively with the PPS with a value of -.37, p (two-tailed) < .01. This correlation contradicts the results from the first phase, where a positive correlation was obtained. The correlations between the SSL and the GPS, product porosity (POR) and PCR variables also experienced a direction change, resulting in large differences in the SSL correlations between the two phases.

From Table 4-7 it can be concluded that the spray variables have the largest correlations with the output variables. The fluidising and seed particle variables show low correlations with most of the variables. The GPS and POR have low correlations with the operating variables, indicating a possible difficulty to develop accurate models for these variables. Most of the correlations are supported from literature; however, there exist some large differences between the correlations obtained from the first phase. An increase in the dataset and experimental levels had a large influence on the significance of the correlations between the input and output variables, as discussed above.

The Spearman results describing the correlations between the different dependent variables are summarised in Table 4-8.

Table 4-8: Third Spearman's rho matrix for phase 2

	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
RUR	1	-.56**	.31*	.19	.18	.34**	-.20	-.40**	-.34**
RER	-.56**	1	-.63**	-.37**	-.11	-.07	.39**	.31*	-.01
EFF	.31*	-.63**	1	.36**	.18	.01	-.53**	-.21	.20
PPS	.19	-.37**	.36**	1	.43**	-.21	-.29*	-.03	-.07
GPS	.18	-.11	.18	.43**	1	-.28*	-.11	.15	.08
POR	.34**	-.07	.01	-.21	-.28*	1	.11	-.17	-.19
PCR	-.20	.39**	-.53**	-.29*	-.11	.11	1	.08	-.33*
PSL	-.40**	.31*	-.21	-.03	.15	-.17	.08	1	.22
GSL	-.34**	-.01	.20	-.07	.08	-.19	-.33*	.22	1

Note: *p<.05, **p<.01

The RUR and RER variables have a negative correlation of -.56, p (two-tailed) < .01. An increase in the RUR will result in a decrease in the RER in order to keep the mass balance checked. The RUR variable has negative correlations with the PSL and GSL variables with values of -.40, p (two-tailed) < .01, and -.34, p (two-tailed) < .01.

All these correlations were obtained in the first phase, where these correlations are due to the influence from the different operating variables. The correlation between the RUR and POR

variable has strengthened to a rho value of .34, p (two-tailed) < .01. An increase in the production rate is mainly a result from an increase in the SLF. The increased spray rates favour the agglomeration mechanism resulting in the production of larger granules with a narrow size distribution. Agglomerates are formed by multiple wetted seed particles colliding and forming solid bridges, resulting in a more porous granule. There exists a large decrease in the correlation values between the RUR and the GPS and PSL variables compared to the screening phase.

The RER and EFF variables have the highest negative correlation in the matrix with a value of -.63, p (two-tailed) < .01. A higher EFF will result in a higher throughput and lower RER. A negative correlation between the RER and PPS with a rho value of -.37, p (two-tailed) < .01, is obtained. A smaller PPS indicates the presence of fines, which will result in an increase in the RER. The correlations between the RER and the GPS and GSL variables observed in the first phase, are not present in the second phase. There is a significant difference in the correlation values between these variables.

The EFF has a moderate negative correlation with the PCR variable, with a correlation value of -.53, p (two-tailed) < .01. An increase in the EFF is due to an increase in either the SLT or SLC variable, as observed from the second matrix. An increase in the SLT or SLC favours the agglomeration mechanism resulting in friable, irregular shaped granules decreasing the PCR. The EFF and PPS have a low positive correlation of .36, p (two-tailed) < .01. Larger granules are produced by the agglomeration mechanism. This correlation was not observed in the first phase. The negative correlation between the EFF and GSL had a value of -.38, p (two-tailed) < .05, in the first phase while the correlation value is .20, with a p-value larger than .05, in the second phase. This indicates a large difference between the correlation values where the correlation became insignificant in the second phase.

The PPS and GPS have the strongest positive correlation in the matrix with a value of .43, p (two-tailed) < .01. The final product stream is obtained by sieving the granulator product stream. The final product stream is a subset of the granulator product stream, resulting in some particle size distribution characteristics present.

There is a large difference in the rho value between the GPS and PSL variables from phase one and two. The first phase showed a moderate negative correlation value of -.42, p (two-tailed) < .05, while in the second phase a correlation value of .15, with a p-value larger than .05, is obtained.

Low correlation values are obtained between the POR variable and the operating and output variables. The POR and GPS variables have a low negative correlation of -.28, p (two-tailed) < .05. The lack of correlations is also present in the first phase, indicating that the POR variable is a complex variable that is influenced by multiple variables at various degrees. It is therefore difficult to obtain high correlation values and subsequently difficult to develop highly accurate models.

The correlation between the PCR and GSL has the largest difference in the correlation values between the two phases. A moderate positive correlation of .47, p (two-tailed) < .01, was obtained in phase one while a moderate negative correlation of -.33, p (two-tailed) < .01, is obtained in the second phase.

Some of the results of the Spearman's rho matrix were found to be the same as in phase one, while there were some correlations with large differences. The increase in modelling data improved the significance of the Spearman's rho correlations from phase one.

4.2.2 Multiple linear regression

Main effects multiple linear regression (MLR-M) and interaction effects (MLR-I) models were developed as discussed in section 3.3.2. The statistical significant coefficients for the main effects and interaction effects MLR models are summarised in Table D-7 and Table D-8 in Appendix D. The statistical parameter results for the models are summarised in Table 4-9.

Table 4-9: Multiple linear regression results for phase 2

	Parameter	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
Main effects (MLR-M)	Independent variables	2	2	2	3			2	3	5
	Model CMD	0.88	0.51	0.61	0.36			0.30	0.37	0.67
	Model ACD	0.87	0.50	0.60	0.32			0.27	0.33	0.64
	Validation CMD	0.94	0.67	0.74	0.48			0.04	0.60	0.77
	Overall CMD	0.90	0.53	0.63	0.33			0.22	0.41	0.68
	SEE	0.43	2.23	6.81	0.13			0.88	0.59	0.40
	MAE	0.30	1.70	5.28	0.09			0.88	0.44	0.29
	MAPE, %	10.4	18.9	8.2	4.8			1.0	8.1	8.3
	Normal distribution	No	No	Yes	Yes			Yes	No	No
Interaction effects (MLR-I)	Independent variables	6	5	3	3	9	13	6	2	5
	Model CMD	0.90	0.73	0.71	0.35	0.43	0.74	0.49	0.36	0.73
	Model ACD	0.89	0.70	0.69	0.31	0.32	0.66	0.43	0.34	0.70
	Validation CMD	0.96	0.85	0.70	0.33	0.66	0.16	0.43	0.60	0.78
	Overall CMD	0.91	0.74	0.71	0.32	0.45	0.62	0.46	0.41	0.74
	SEE	0.40	1.71	5.88	0.13	0.10	1.56	1.20	0.58	0.36
	MAE	0.28	1.31	4.06	0.09	0.07	1.05	0.80	0.42	0.27
	MAPE, %	9.1	14.7	6.2	4.8	4.8	2.8	0.9	7.8	7.8
	Normal distribution	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes

The production rate (RUR) main effects and interaction effects models performed the best fitting 90% of the data in a near linear relation as shown in Figure 4-10 and Figure 4-11. The MLR-M model fit the validation data very well, with the model regression line near the perfect linear regression line (y=x) as shown in Figure 4-10. The interaction effects RUR model performed almost identical to the main effects model. The MLR-I included 4 additional independent variables which contributed significantly to the interaction model, since there is only a 1% difference between the coefficient of multiple determination (CMD) value and the adjusted coefficient of multiple determination (ACD) value.

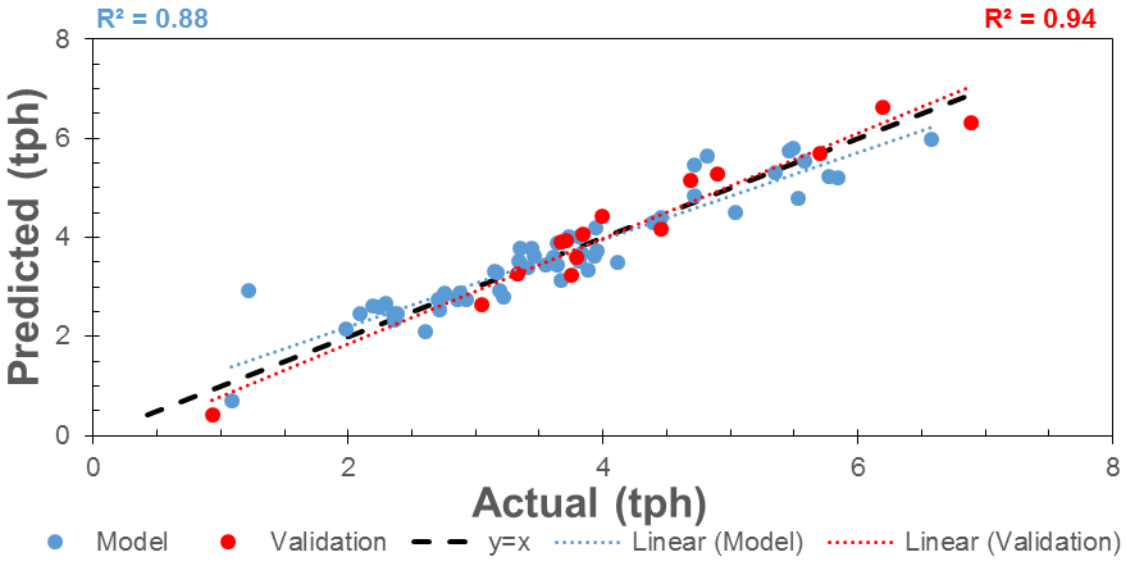


Figure 4-10: Phase 2 MLR-M RUR model regression

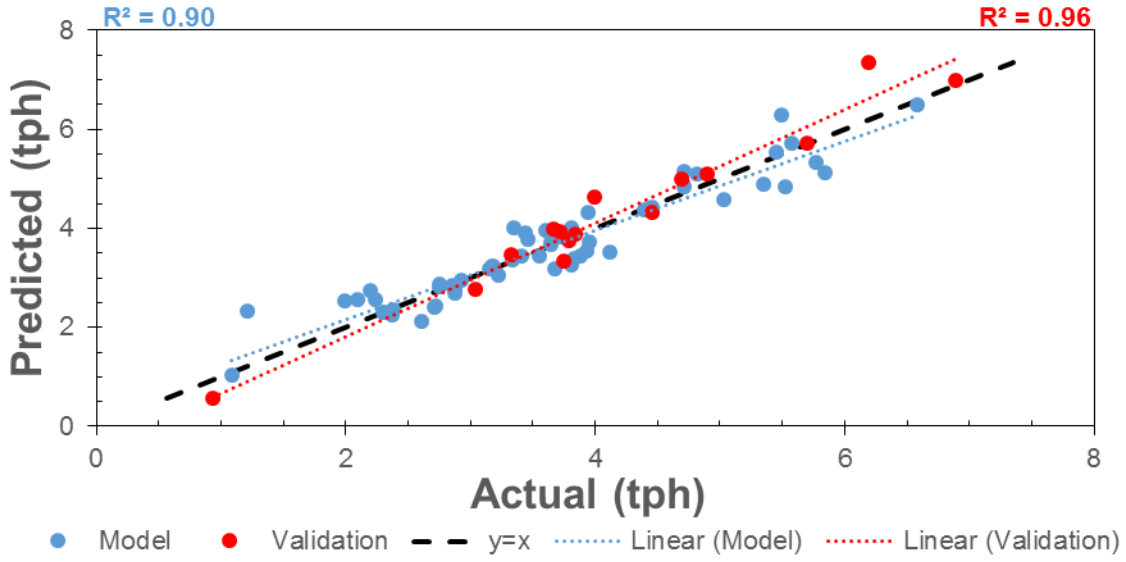


Figure 4-11: Phase 2 MLR-I RUR model regression

The interaction effects RUR model has a lower standard error of estimate (SEE), resulting in a narrower distribution of the data points as shown in Figure 4-11. The mean absolute error (MAE) for the interaction effects model is 0.02 lower than the main effects RUR model indicating a small increase in accuracy with the interaction effects included. Both models have a moderate mean absolute percentage error (MAPE) value due to a few large deviations in some data points. These data points are predicted at a higher RUR values compared to the actual values, resulting in a large positive error. Hyndman & Koehler (2006) stated the MAPE penalises positive errors heavier, resulting in the higher MAPE values for the RUR models.

Including the interaction variables resulted in the normal distribution of the modelled data, resulting in the confirmation of the assumption made for the MLR modelling.

The MLR-I model for the recycle ratio (RER) performed better than the MLR-M model, with an increase of 20% in the model's CMD. The additional interaction variables increased the performance of the model with only a small decrease of 3% in the ACD value. The RER interaction effects model has a moderate performance, fitting 85% of the validation data but with a moderately high MAPE value of 15%. The interaction effects model's SEE and MAE values decreased by 0.5 and 0.4 respectively compared to the main effects model. The MLR-M and MLR-I models' accuracies of the model and validation data are shown in Figure 4-12 and Figure 4-13.

The main effects RER model exhibit large errors on the modelling data, especially at data points 30 to 43. The MLR-I model shows an increase in fitting the data trend, as seen in Figure 4-13 at data points 43 to 58. The model still has a low accuracy and miss predicts data points 35 to 38 just as the MLR-M model. The large deviations in these points result in the large MAPE values obtained.

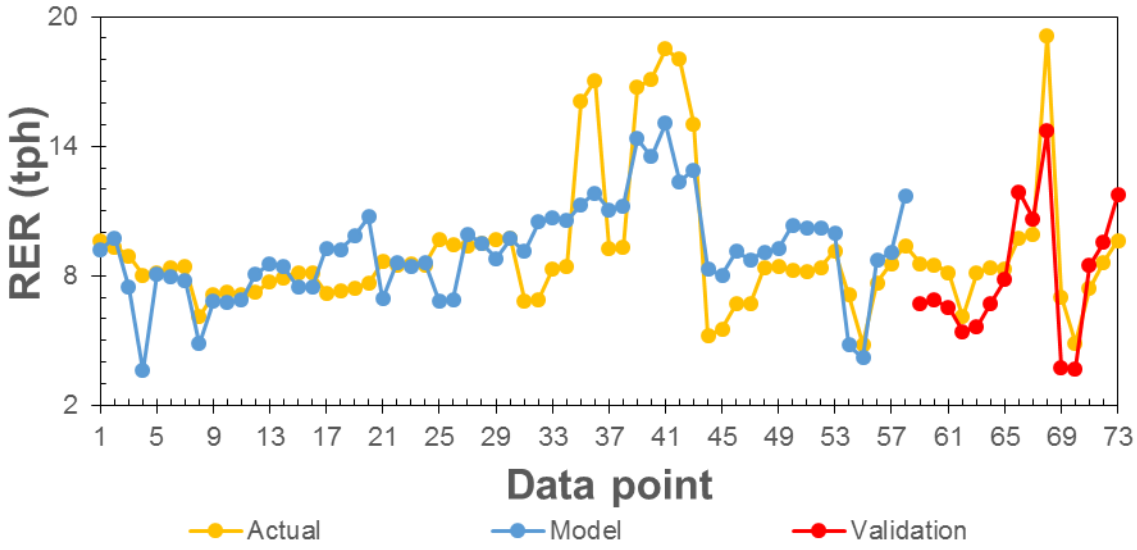


Figure 4-12: Phase 2 MLR-M RER model performance

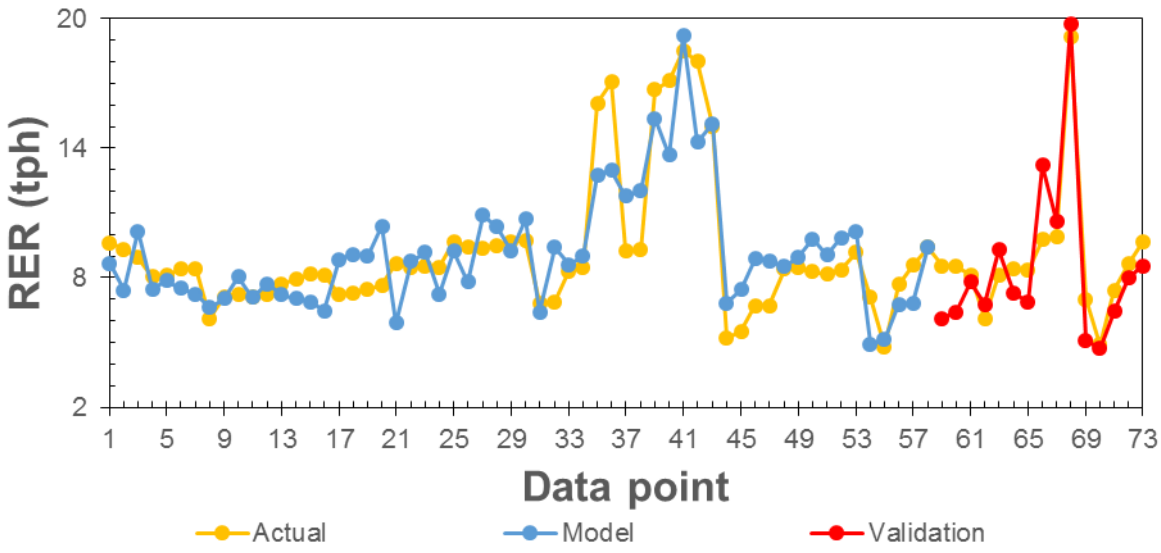


Figure 4-13: Phase 2 MLR-I RER model performance

There are a few data points that are outside the mean, e.g. data points 35, 36, 39 to 43 and 68. These data points account for about 10% of the data, resulting in difficulty to accurately fit these points. Both models performed well on these points, with the MLR-I model fitting some of these points with a high accuracy, especially on the high validation point at 68. The MLR-I model has a higher accuracy with a lower errors compared to the MLR-M model. The addition of the interaction effects terms also resulted in a normal distribution of the residuals for the RER MLR model.

The efficiency (EFF) interaction model also outperformed its main effects counterpart with an increase of 10% of the model data CMD. The MLR-I model predicts the validation data reasonably well with a 30% regression variation. The EFF models have the largest SEE values, indicating a large deviation in the model's prediction values. The MLR-I model has a lower MAE value compared to the MLR-M model, indicating that the MLR-I model has a higher prediction accuracy. The large error values are observed from the performance graphs shown in Figure 4-14 and Figure 4-15.

The EFF main effects model shows large deviations between the actual and predicted values, especially at the extreme EFF values, resulting in the large SEE and MAE values. The MLR-M model does not follow the general trend of the data well, as seen at data points 17 to 20 and 25 to 26. The MLR-I model fits the data points more accurately compared to the MLR-M model, but still underperforms on data points 25, 36 and 42 to 47. The MLR-I model follows the general trend of the data and fits the average validation values well. The MLR-I model is concluded to have a moderate performance with a CMD value of 71% and a fair accuracy with a MAPE value of 6.2%.

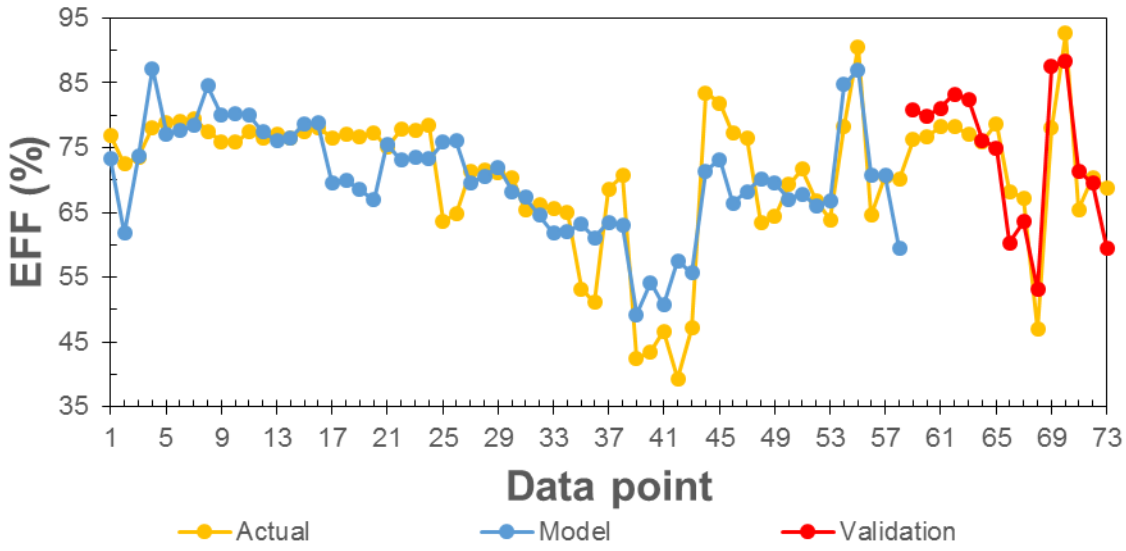


Figure 4-14: Phase 2 MLR-M EFF model performance

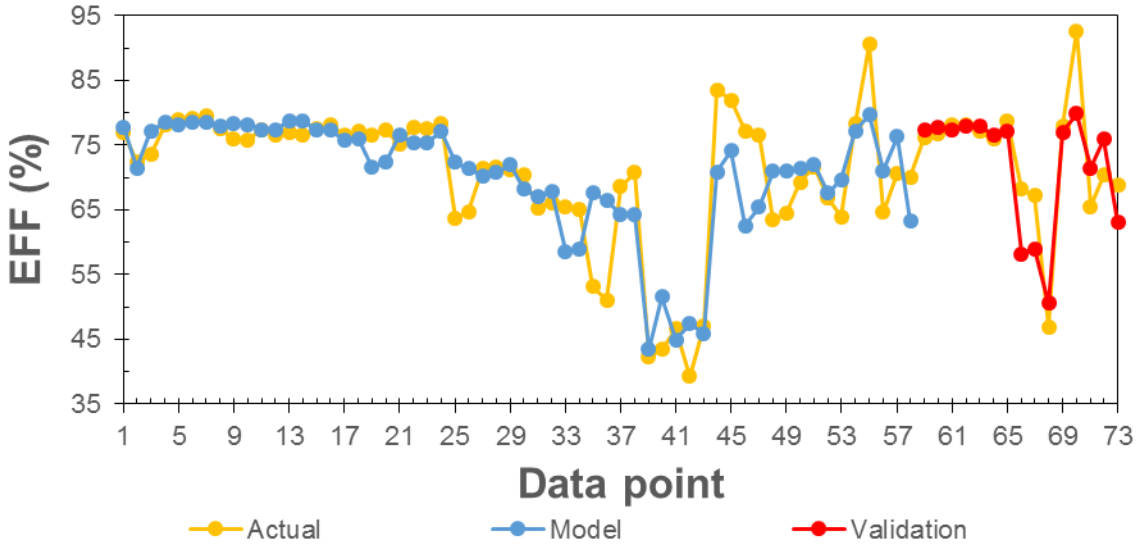


Figure 4-15: Phase 2 MLR-I EFF model performance

The MLR-M model for the product mean particle size (PPS) performed slightly better than the MLR-I model with a 15% higher validation regression value. Both models have low CMD values indicating a low prediction performance from both models. The moderate MAPE values also indicate a low accuracy performance from the models as shown in Figure 4-16 and Figure 4-17. The low SEE and MAE values are due to the low range of the PPS values, making the error values difficult to trust and interpret. Both models follow the general trend of the data fairly, with the exception of a few points. The MLR-M model fits the data points with a higher accurately and follows the general trend better, but it is still considered inaccurate for prediction based on the statistical and visual evaluation.

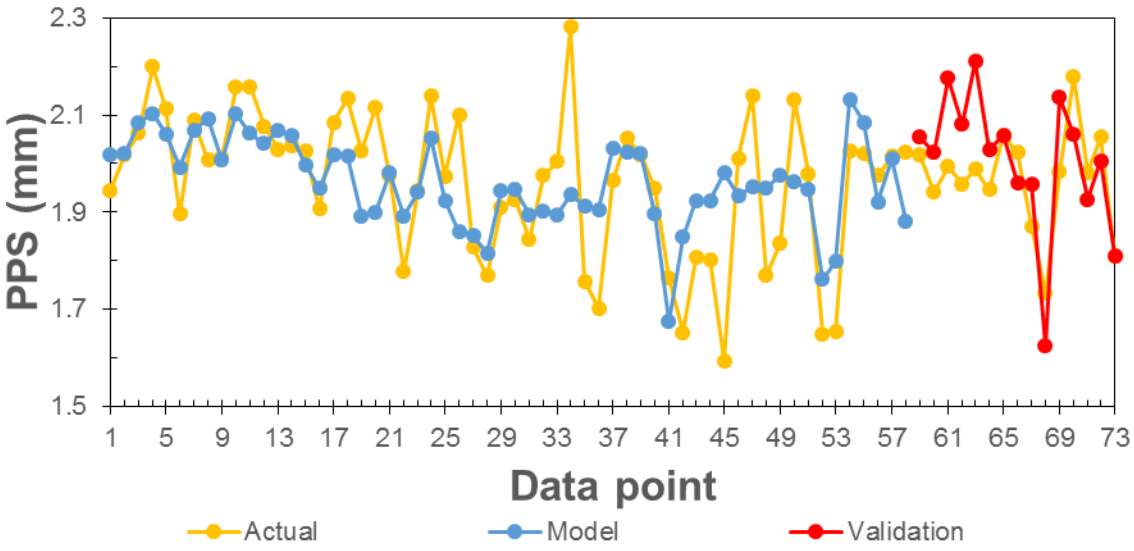


Figure 4-16: Phase 2 MLR-M PPS model performance

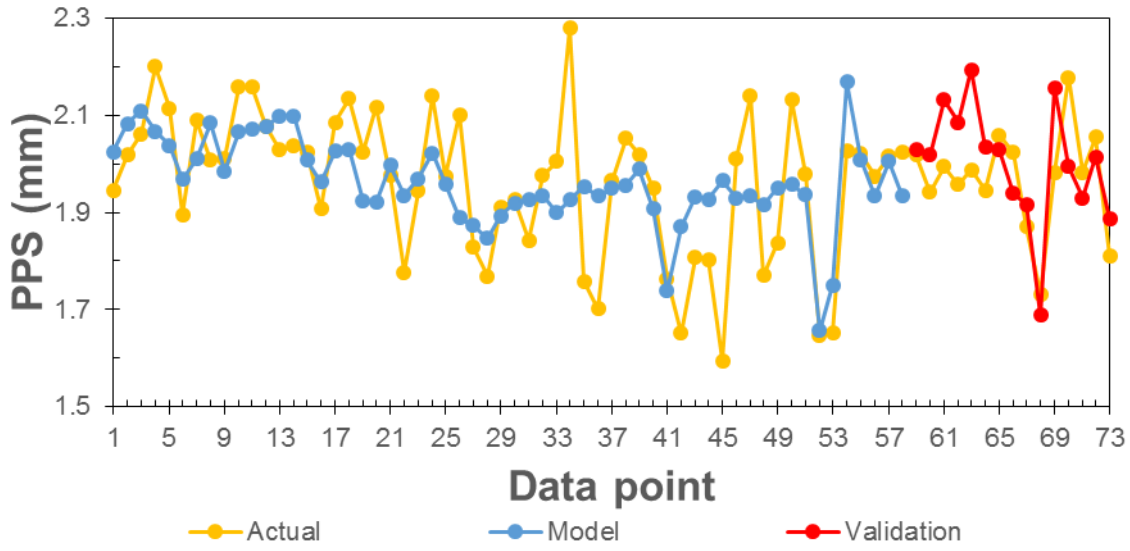


Figure 4-17: Phase 2 MLR-I PPS model performance

Main effects statistical significant models could not be developed for the granulator mean particle size (GPS) and the product porosity (POR). The GPS interaction effects model did not perform well with an 11% difference between the CMD and the ACD values, as observed in Table 4-9. This suggests that some independent variables do not significantly increase the model's performance and could be excluded. Eliminating these variables will result in a decrease in the CMD of the model. The low SEE and MAE values are due to the low values from the GPS variable, resulting in inaccurate values. The moderate MAPE value indicates a low accuracy of the GPS model.

The interaction effects POR model performed well with a 74% model CMD, however, there is an 8% decrease between the model CMD and the ACD values. The validation regression is very low with a CMD value of 16%, indicating a weak validation performance from the POR model as seen in Figure 4-18. The POR model has a low MAPE value of 3% indicating a good accuracy on the predicted values. The POR model follows the trend reasonably with a fair accuracy on the modelling points; however, the model performs very poorly on the validation data, rendering this model inaccurate for prediction purposes.

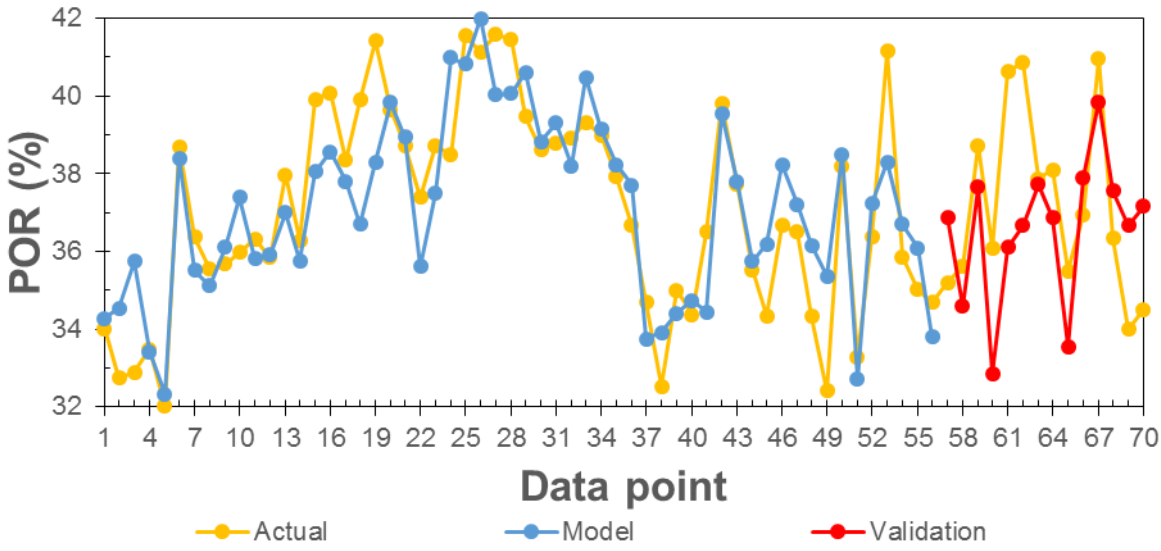


Figure 4-18: Phase 2 MLR-I POR model performance

The product circularity (PCR) interaction model has a 19% improvement in the model's regression fit with a 40% increase in the validation regression fit compared to the main effects model. The SEE is 0.4 higher than the MLR-M model indicating a larger deviation but the model has a lower MAE specifying smaller error differences. The larger SEE is due to the increase in the number of independent variables resulting in the calculation of a higher SEE. The MAPE for both models are low, 1%, indicating a lower normalised error between the values. Both models do not fit the modelling or validation data with a high accuracy as shown in Figure 4-19 and Figure 4-20. Some data points are lower than the mean, e.g. data points 51 to 52, 57 to 58 and 72 to 73. The models struggle to predict these lower values resulting in high error values and reducing the CMD values. Both models follow the trend of the data reasonably, but with a low accuracy. The MLR-I model follows the validation trend better compared to the MLR-M model, but with large errors. Both models are concluded inappropriate for prediction purposes due to the low accuracies.

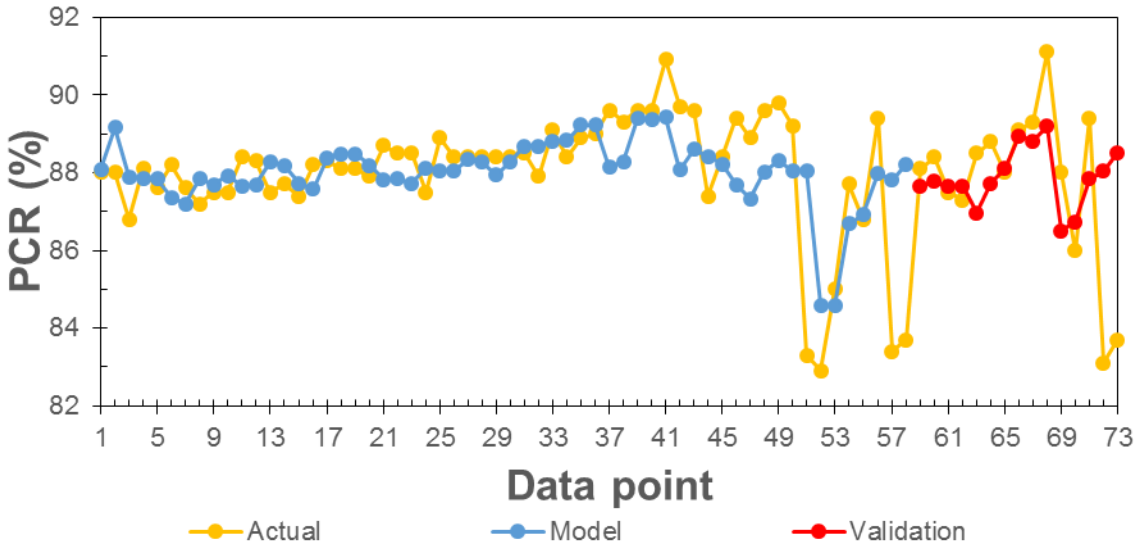


Figure 4-19: Phase 2 MLR-M PCR model performance

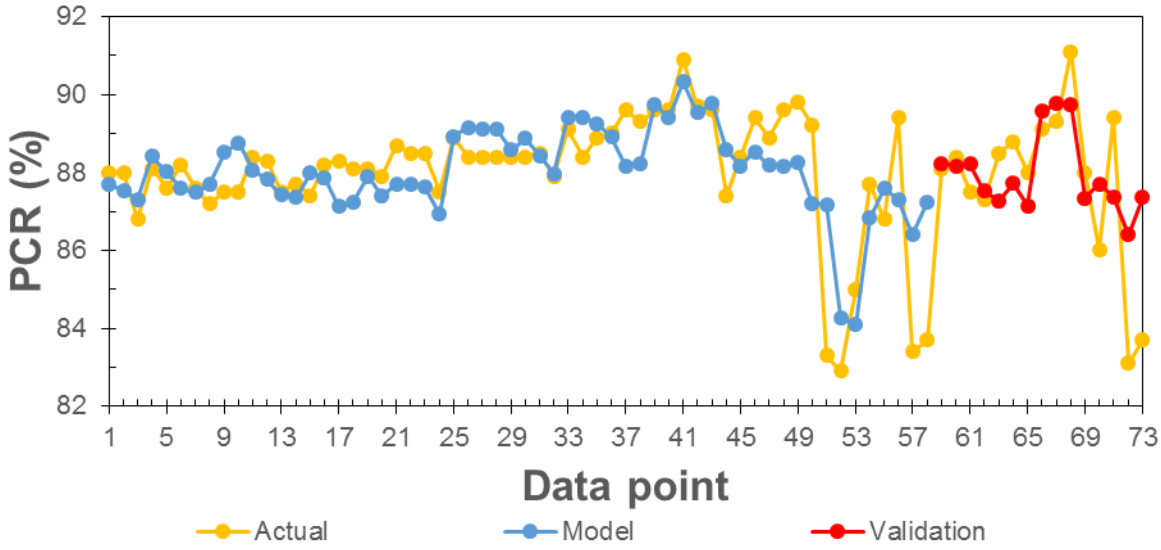


Figure 4-20: Phase 2 MLR-I PCR model performance

The main effects and interaction effects product particle slope (PSL) models have almost identical performances. The MLR-I model has a 1% lower model regression fit and only contains two statistical significant variables compared to the three from the MLR-M model, resulting in a smaller difference between the CMD and the ACD values for the MLR-I model. The MLR-I model has a lower SEE and MAE values that are due to the smaller number of independent variables. Both models follow the general trend of the data but with low accuracies as observed in Figure 4-21 and Figure 4-22. Both models struggle to predict the high extreme values at points 13, 41 and 68. The errors from these points contributed to the moderate MAPE value. The MLR-I model follows the validation data with a higher accuracy compared to the MLR-M model; however, both models are impractical to use due to the low model accuracies.

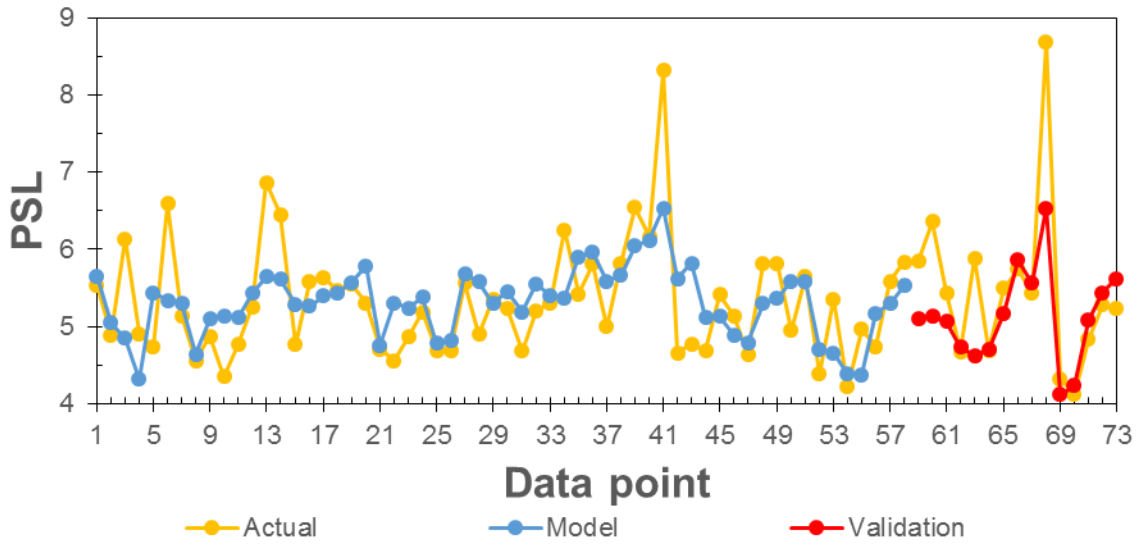


Figure 4-21: Phase 2 MLR-M PSL model performance

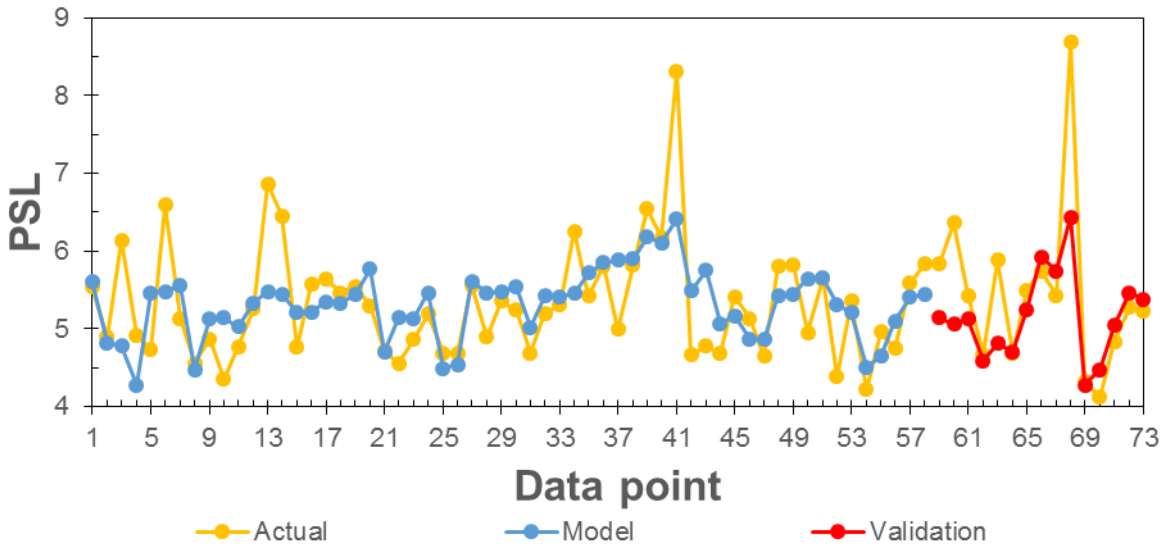


Figure 4-22: Phase 2 MLR-I PSL model performance

The granulator particle slope (GSL) interaction effects model performed with a higher CMD values compared to the main effects model. The low deviations in the predicted values and almost similar performances from both models can be seen in Figure 4-23 and Figure 4-24. The MLR-I model predicts with a higher accuracy as seen on the high data points at 16 and 41, resulting in the lower error values. The MLR-I model follows the model and validation data trend better, giving higher accurate predictions. There are still a few high and low data points that the model struggles to predict, resulting in the moderate MAPE value. The MLR-I model has an overall higher accuracy with lower error values compared to the MLR-M model, making it the better performing model between the two.

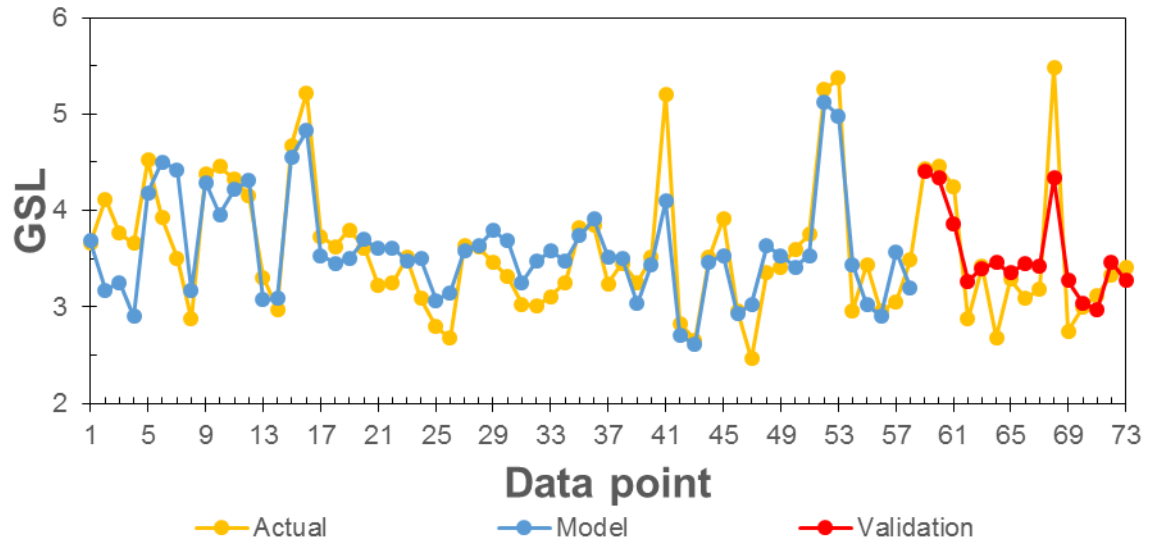


Figure 4-23: Phase 2 MLR-M GSL model performance

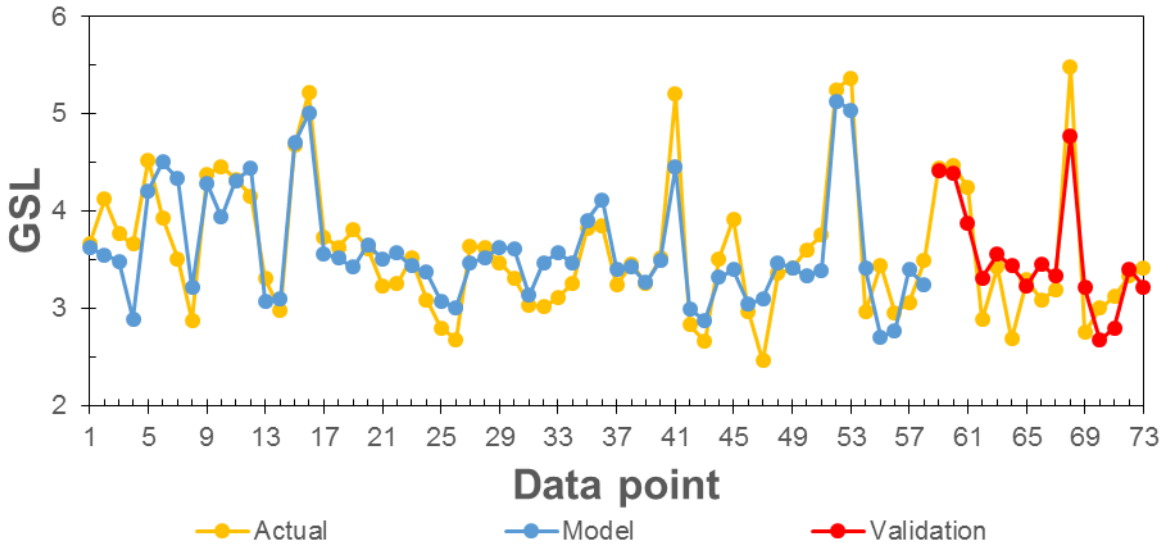


Figure 4-24: Phase 2 MLR-I GSL model performance

It is concluded that the RUR, RER, EFF and the GSL models performed well with moderately high accuracies and low errors. The MLR-I models performed in general better than the MLR-M models, indicating that two-way interactions are important to take into account in this process. Some of the main effects models also performed well especially the RUR and PPS models. The PPS, GPS, POR, PCR and PSL models did not perform well overall but had moderate performances with low accuracies based on their respected regression and error values. The PPS and PSL models yielded very low model CMD values indicating the difficulty to accurately fit the data. The residuals of most models were normally distributed, verifying the modelling assumption for MLR technique.

4.2.3 Comparison between MLR and ANN models

The interaction effects multiple linear regression (MLR-I) models discussed in section 4.2.2 are compared with artificial neural network (ANN-I) models including all the independent variables in order to evaluate the two modelling techniques. The results for both models are summarised in Table 4-10.

Table 4-10: Phase 2 MLR model comparison with the ANN models

	Parameter	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
Multiple linear regression	Independent variables	6	5	3	3	9	13	6	2	5
	Model CMD	0.90	0.73	0.71	0.35	0.43	0.74	0.49	0.36	0.73
	Model ACD	0.89	0.70	0.69	0.31	0.32	0.66	0.43	0.34	0.70
	Validation CMD	0.96	0.85	0.70	0.33	0.66	0.16	0.43	0.60	0.78
	Overall CMD	0.91	0.74	0.71	0.32	0.45	0.62	0.46	0.41	0.74
	SEE	0.40	1.71	5.88	0.13	0.10	1.56	1.20	0.58	0.36
	MAE	0.28	1.31	4.06	0.09	0.07	1.05	0.80	0.42	0.27
	MAPE, %	9.1	14.7	6.2	4.8	4.8	2.8	0.9	7.8	7.8
	Normal distribution	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes
Artificial neural network	Independent variables	7	7	7	7	7	7	7	7	7
	Model CMD	0.92	0.85	0.76	0.36	0.27	0.31	0.61	0.59	0.78
	Model ACD	0.91	0.82	0.72	0.27	0.17	0.21	0.55	0.54	0.75
	Validation CMD	0.94	0.80	0.51	0.55	0.23	0.20	0.66	0.72	0.89
	Overall CMD	0.93	0.83	0.67	0.36	0.26	0.29	0.61	0.64	0.80
	SEE	0.39	1.32	5.65	0.13	0.11	2.33	1.06	0.49	0.33
	MAE	0.28	0.98	4.08	0.10	0.08	1.79	0.75	0.37	0.26
	MAPE, %	8.6	11.8	6.1	4.9	5.1	4.9	0.9	6.9	7.6
	Best epochs	17	5	4	8	16	6	19	15	13

The production rate (RUR) ANN-I model performed slightly better with higher coefficient of multiple determination (CMD) values for the model, validation and overall datasets. The additional variable in the ANN-I model does not influence the adjusted coefficient of multiple determination (ACD) value, due to the large model regression value. The standard error of estimate (SEE) and mean absolute error (MAE) values are similar for both models. The low SEE values indicate a low prediction deviation from the actual data as shown in Figure 4-25 and Figure 4-26. Moderate mean absolute percentage error (MAPE) values are obtained for both models due to a few data points that were predicted with a higher RUR value compared to the actual data. These data points resulted in positive error values that are penalised more by the MAPE parameter (Hyndman & Koehler 2006). The large MAPE is significantly influenced by data point 40, where a large positive error is obtained at a low actual RUR value of 1.09 tph.

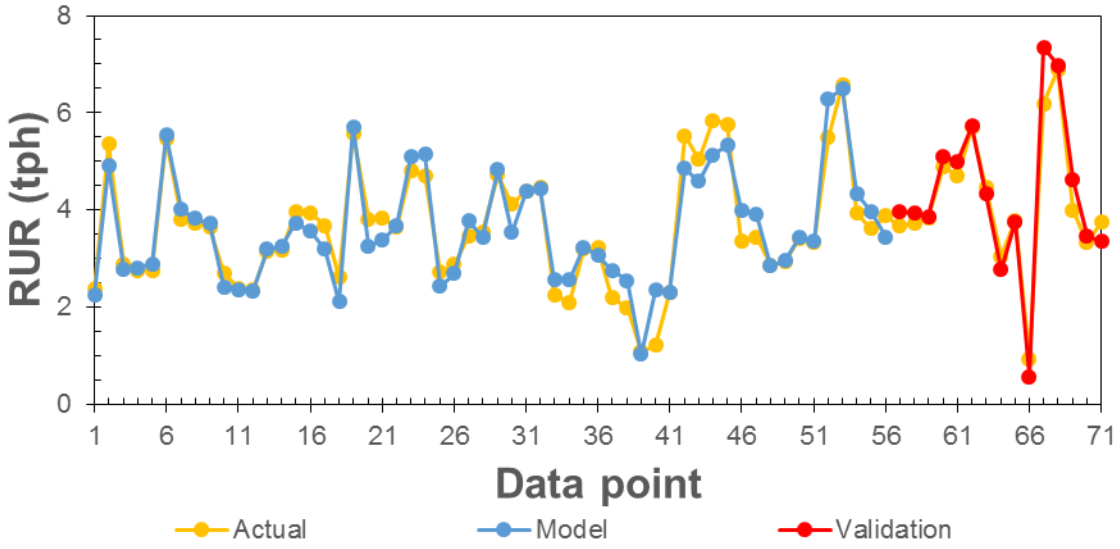


Figure 4-25: Phase 2 MLR-I RUR model performance

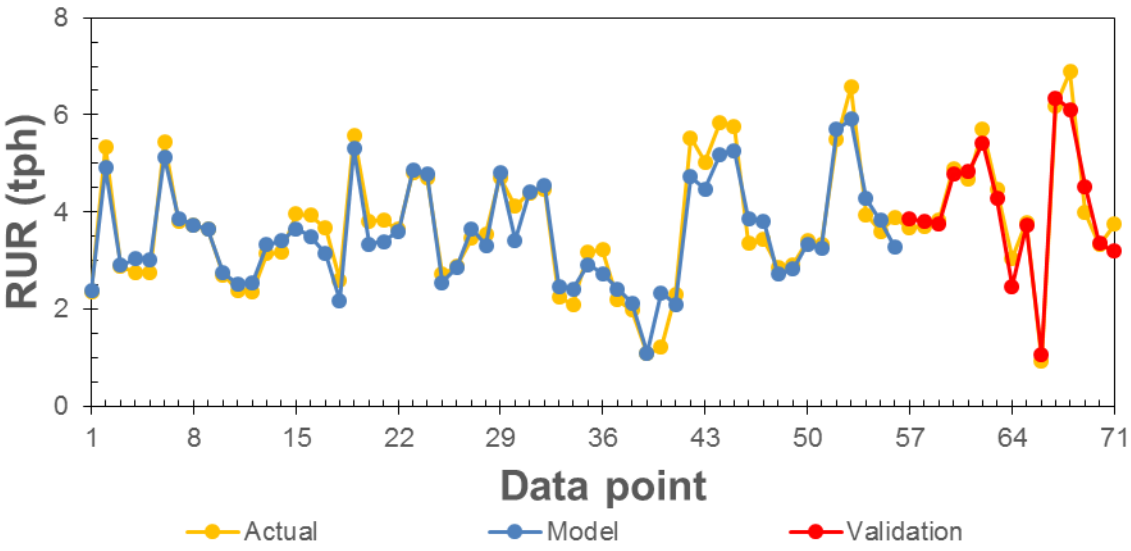


Figure 4-26: Phase 2 ANN-I RUR model performance

The RUR ANN-I and MLR-I models appear to fit and follow the data trend very well with a high level of accuracy. The MLR-I model fits a few modelling data points with a higher accuracy in comparison to the ANN-I model, especially at data points 3 to 5, 35 and 36. The MLR-I model also fits the validation points with a higher accuracy compared to the ANN-I model, except for data points 67 and 71. The regression values for the ANN-I model are closer to each other, indicating a more reliable prediction performance. Both models exhibit a good performance with high accuracies, making them both ideal for predicting the RUR values.

The recycle ratio (RER) MLR-I model has a 12% lower model CMD but with a 5% higher validation CMD compared to the ANN model. The ANN-I model has lower error values indicating a lower prediction deviation from the actual data. The ANN-I model's regression values are higher with the data being distributed closer to the actual points as shown in Figure 4-28. The validation data for the RER model is distributed with only a single validation point at the extreme end of the range. Both models predict this point with a high level of accuracy. This data point is seen on the far right-hand corner of Figure 4-27 and Figure 4-28.

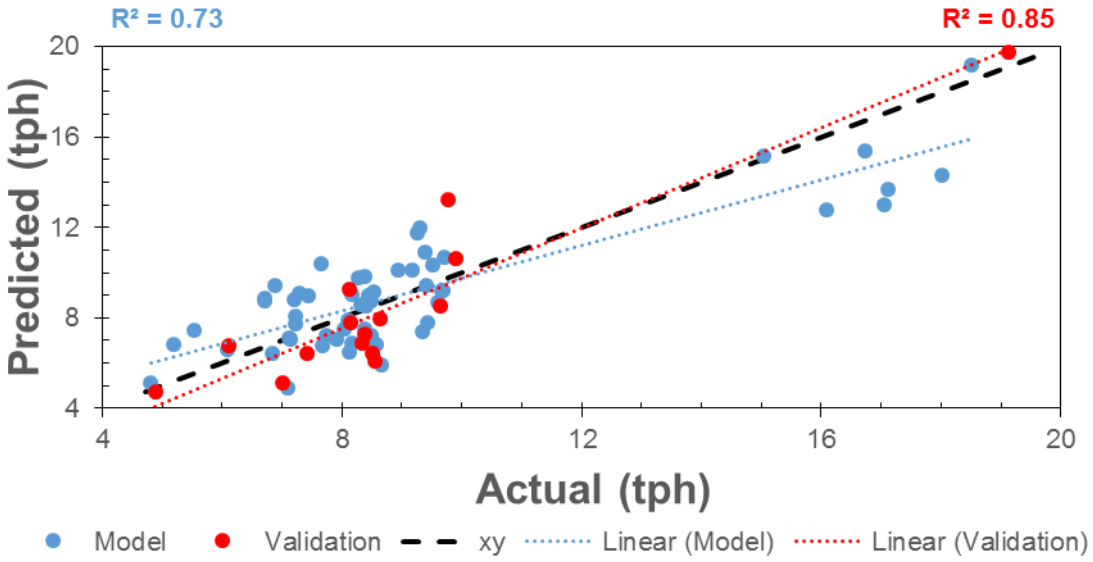


Figure 4-27: Phase 2 MLR-I RER model regression

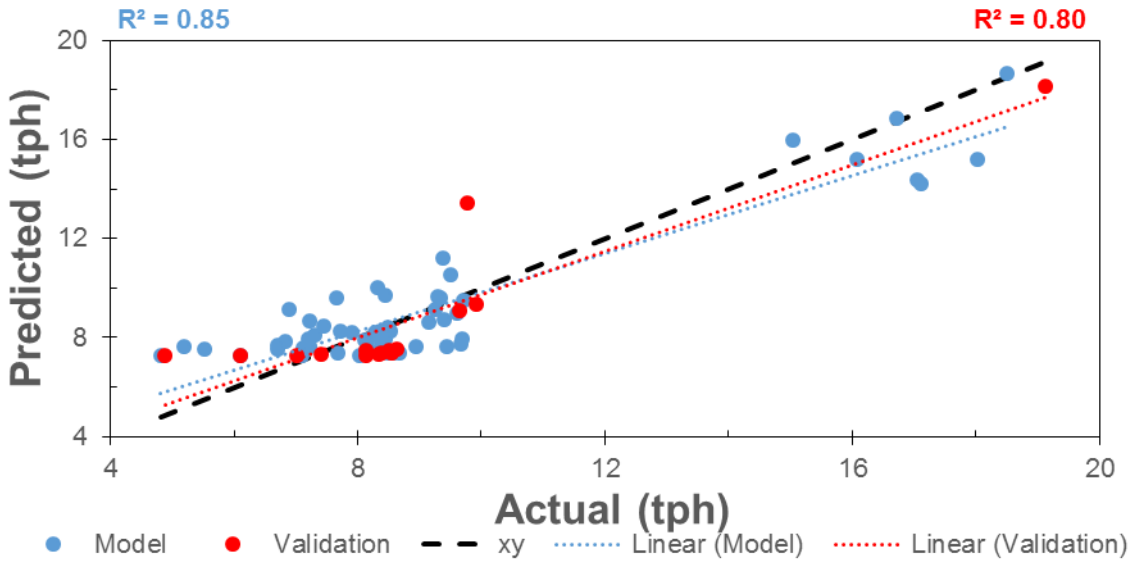


Figure 4-28: Phase 2 ANN-I RER model regression

The RER MLR-I model struggles to fit the modelling data points far from the mean with a high accuracy, resulting in scattered points below the perfect regression line. The mean modelling data points are also scattered above and around the perfect regression line resulting in a low model regression fit. The moderately large MAPE values are due to the large deviations between the predicted and actual values as shown on these figures. The ANN-I model fits more data points with a higher accuracy, resulting in a smaller MAPE value. Both regression lines have almost similar slopes indicating a high prediction reliability between the modelling and validation data. The ANN-I model performed with a higher accuracy compared to the MLR-I model. Both models performed well within an acceptable accuracy, making both models usable for prediction purposes.

The ANN-I efficiency (EFF) model has a larger model regression fit but with a 4% difference in the CMD and ACD values, indicating that the additional variables do not improve the models performance significantly. The MLR model has a 20% higher validation fit with a slightly higher SEE and MAE value. Both models experienced difficulty to accurately predict the EFF values, resulting in large SSE values indicating a large deviation in the predicted results as shown in Figure 4-29 and Figure 4-30. The MLR model followed the data trend with a higher accuracy, but shows large error values as observed at data points 36, 44 to 47 and 55. Some EFF values appears to be distributed within a data range, e.g. between an EFF value of 70% and 85%, while only a few values crossing this range. This might be a reason for the weak performances at the extreme points for both models.

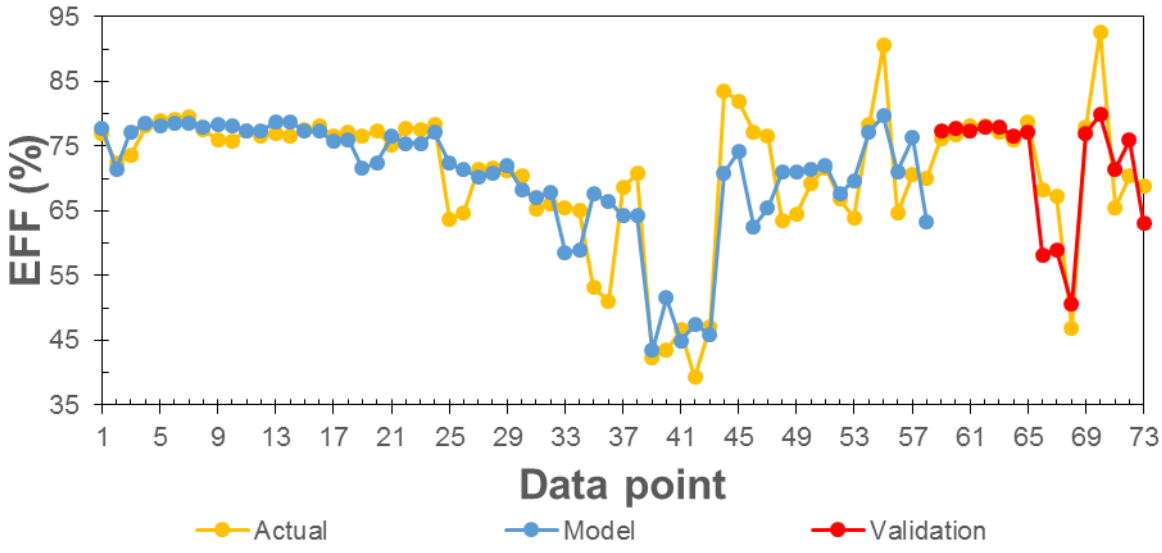


Figure 4-29: Phase 2 MLR-I EFF model performance

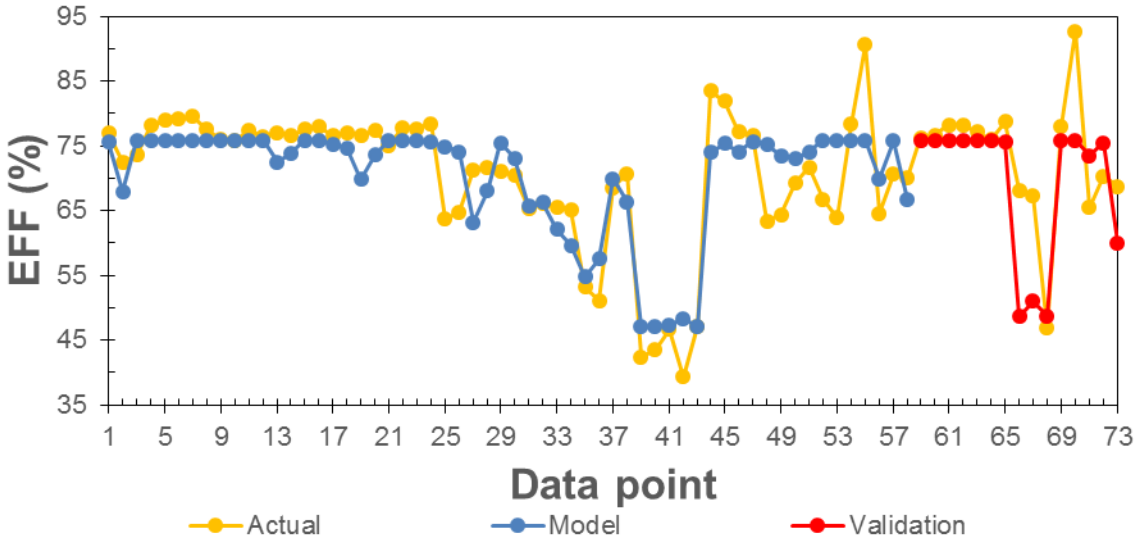


Figure 4-30: Phase 2 ANN-I EFF model performance

The ANN-I EFF model follows the general trend but with a low accuracy. The ANN-I model predicts some data points with a constant value, indicating undertraining. The EFF ANN-I model has the lowest training epoch value in the table. The validation dataset contains some extreme EFF values that could cause the validation error to increase with an increase in training. The MLR model performed better than the ANN-I model, having all its CMD values close to 70%, indicating a more reliable prediction performance.

The ANN-I product mean particle size (PPS) model has a 22% higher validation regression fit compared to the MLR-I model, but with a 9% difference between the model CMD and ACD values. Both models follow the general tendency with a low accuracy, underperforming on the extreme points as presented in Figure 4-31 and Figure 4-32.

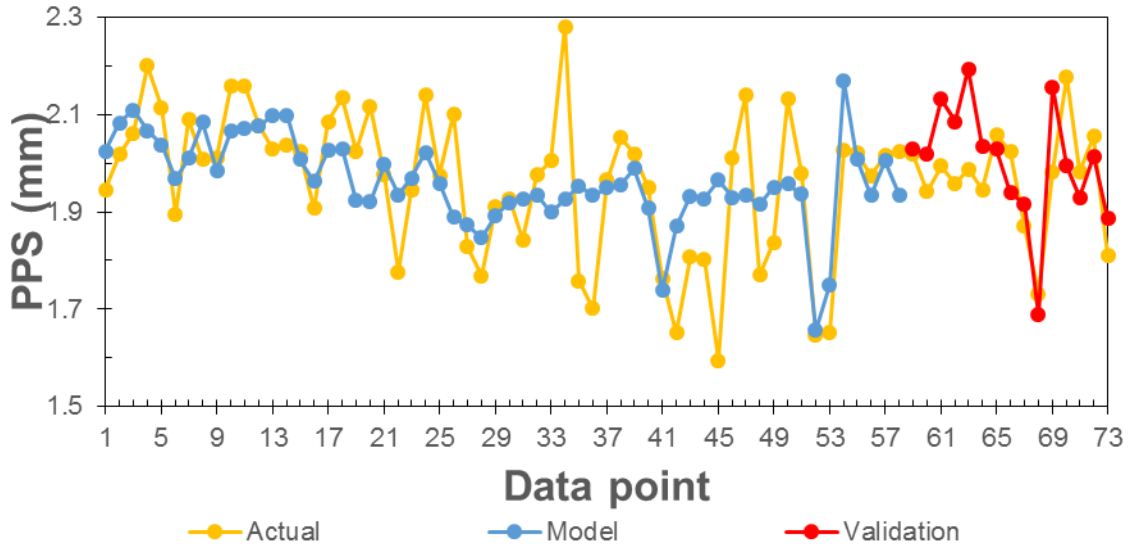


Figure 4-31: Phase 2 MLR-I PPS model performance

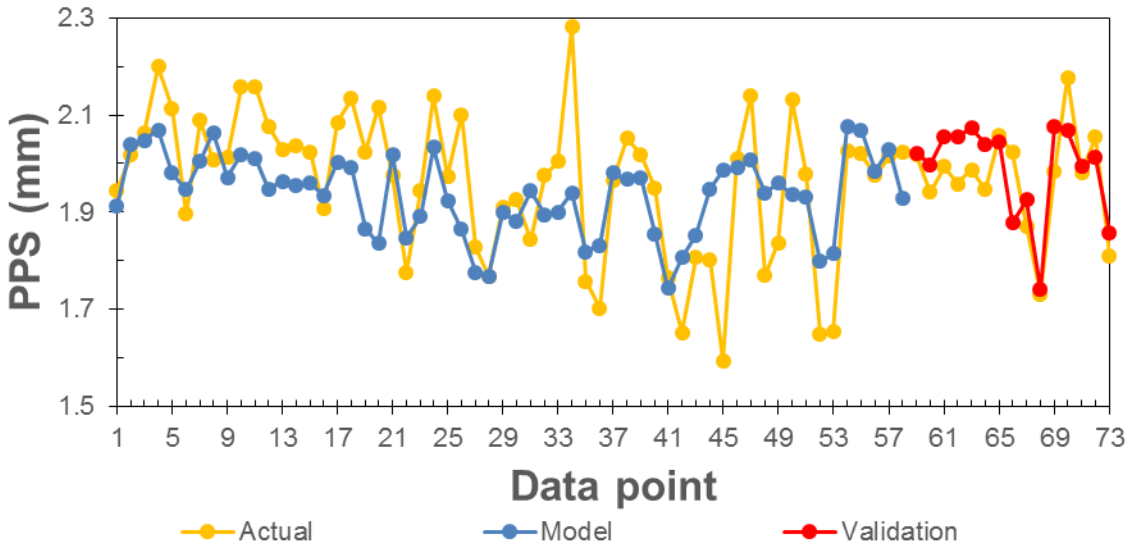


Figure 4-32: Phase 2 ANN-I PPS model performance

Both models have almost identical error low error values. These low values are due to the range of the PPS values. The MAPE value is moderate, indicating a significant difference between the actual and predicted values. The ANN-I model follows the trend with a higher accuracy, but shows relatively large deviations. The CMD values for the MLR-I model are closer to each other, indicating a more reliable prediction performance compared to the ANN-I model. Both models have difficulty to provide accurate predictions and therefore cannot be used for prediction purposes.

The MLR-I granulator mean particle size (GPS) model outperformed the ANN-I model with a 16% higher model regression fit and a 40% higher validation regression fit compared to the ANN-I model. Both models have a 10% difference between their CMD and ACD values, indicating that some variables do not contribute significantly to the models. The regression plots of both models are shown in Figure 4-33 and Figure 4-34. The model prediction values for the MLR-I model are distributed around the perfect regression line. There are some model and validation data points that are predicted away from the actual values. The ANN-I GPS model exhibits a near horizontal regression fit, indicating a lack in the prediction capability of the model. This is an indication of insufficient training, despite the large number of training epochs. The ANN-I model has larger error values due to its weaker prediction. The MLR-I model has a higher accuracy and better performance but finds it challenging to predict the values with a high accuracy, making both models inaccurate for prediction.

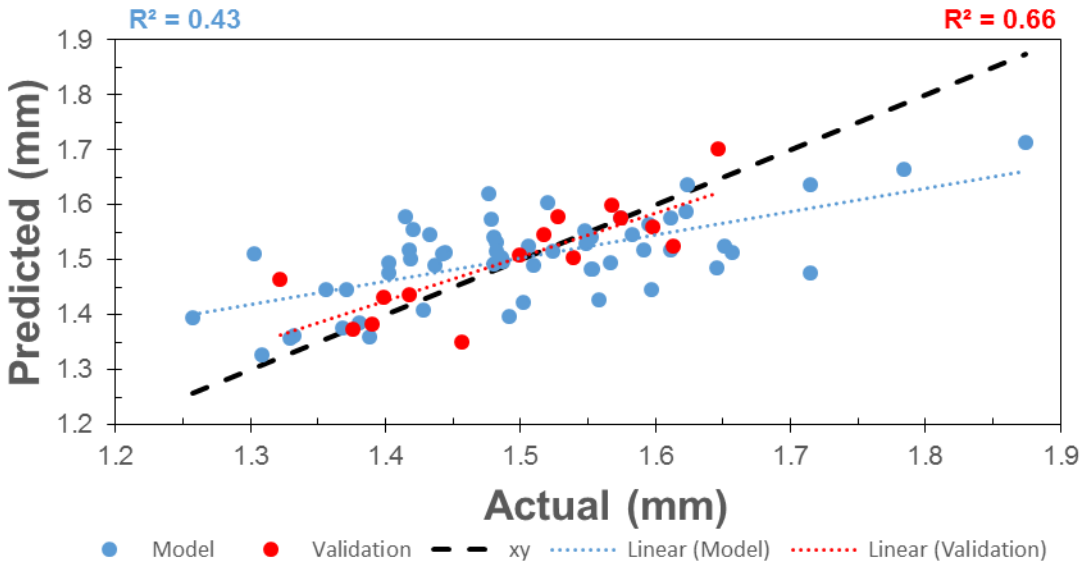


Figure 4-33: Phase 2 MLR-I GPS model regression

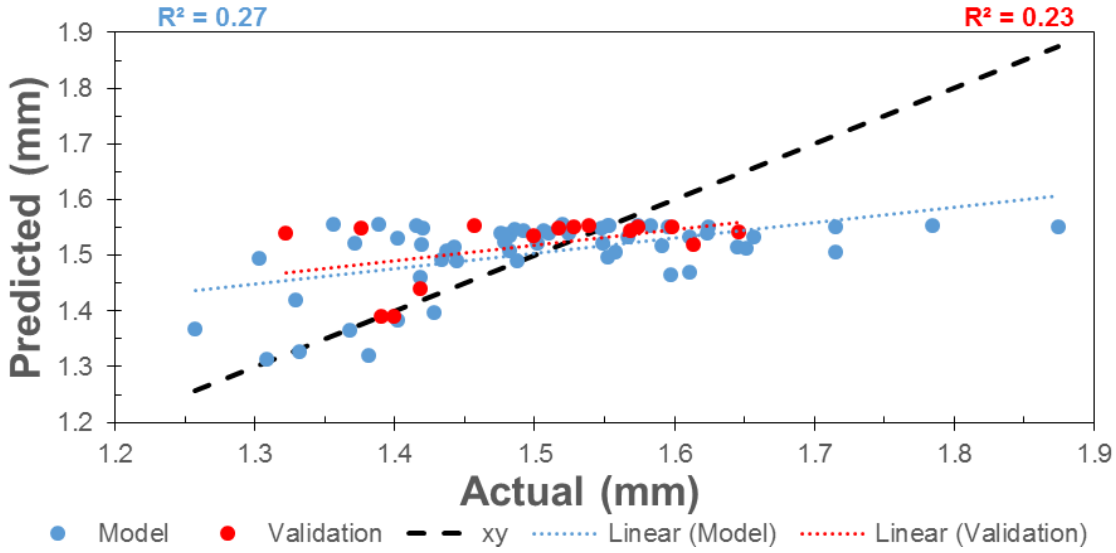


Figure 4-34: Phase 2 ANN-I GPS model regression

The MLR-I product porosity (POR) model has a high model regression fit but with a low validation regression fit compared to the ANN-I POR model. Including additional variables influenced both models greatly with a 10% difference in the CMD and ACD values. This indicates than some variables are unnecessary and can cause instability in the models. The MLR-I model fits the data more accurately, resulting in a close distribution of data points around the perfect regression line, as shown in Figure 4-35. The ANN-I model's CMD value is low due to the validation check stopping the training proses too early, resulting in insufficient training as indicated by the near horizontal regression fit in Figure 4-36. The ANN-I model has higher SEE and MAE values due to the low model fit. The MLR-I model performs better than the ANN-I model, but has a low validation performance with large deviations concluding that both models are not suited for prediction.

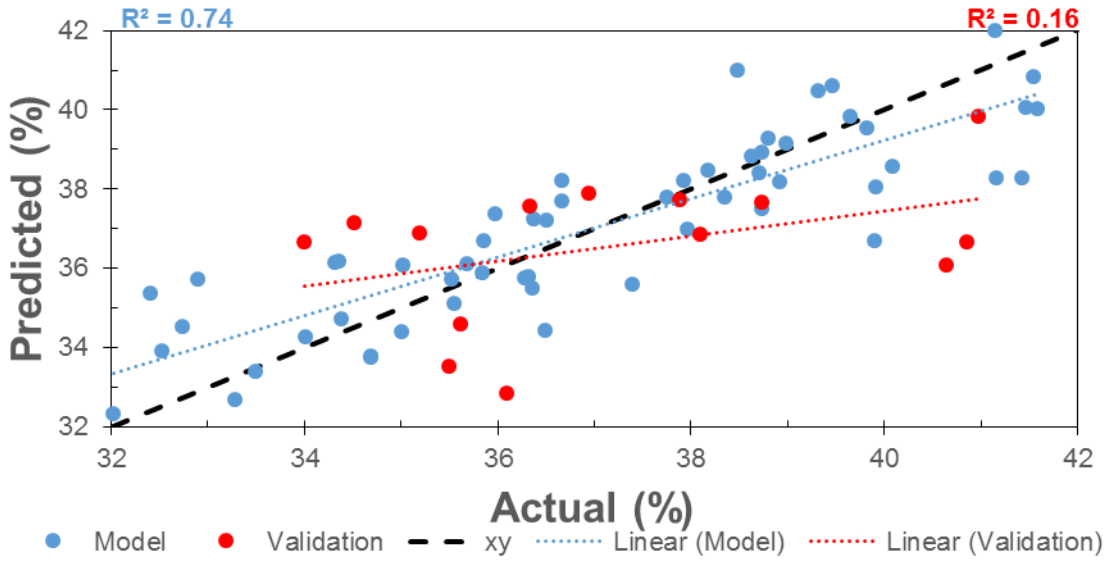


Figure 4-35: Phase 2 MLR-I POR model regression

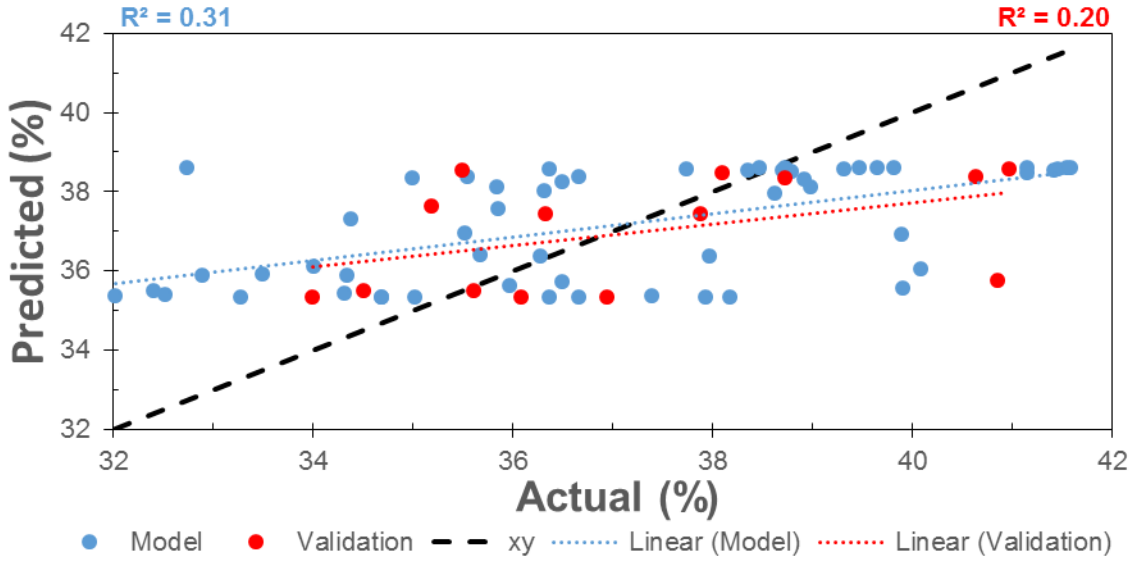


Figure 4-36: Phase 2 ANN-I POR model regression

The ANN-I product circularity (PCR) model has larger modelling, validation and overall CMD values. The performance of the ANN-I model is lower compared to the MLR-I model as shown in Figure 4-37 and Figure 4-38. The MLR-I model appears to follow the data trend with a moderate accuracy. There are some points that differ largely from the actual values resulting in the moderate SEE values. The MLR-I model shows lower differences between the CMD and ACD values compared to the ANN-I model. The MLR-I model finds it difficult to predict the validation points with a high accuracy, especially data points 64, 72 and 73.

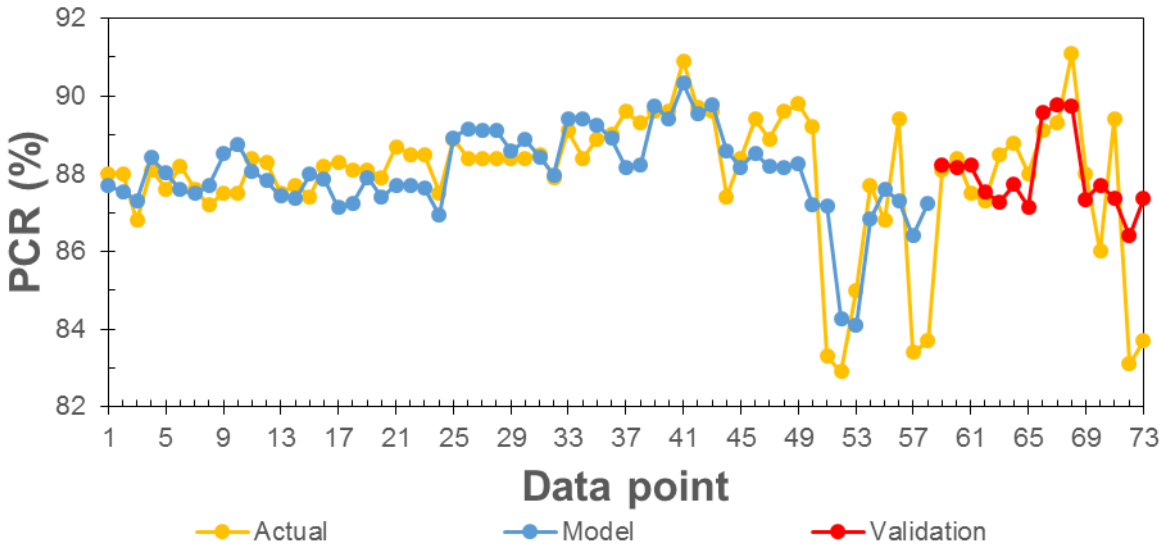


Figure 4-37: Phase 2 MLR-I PCR model performance

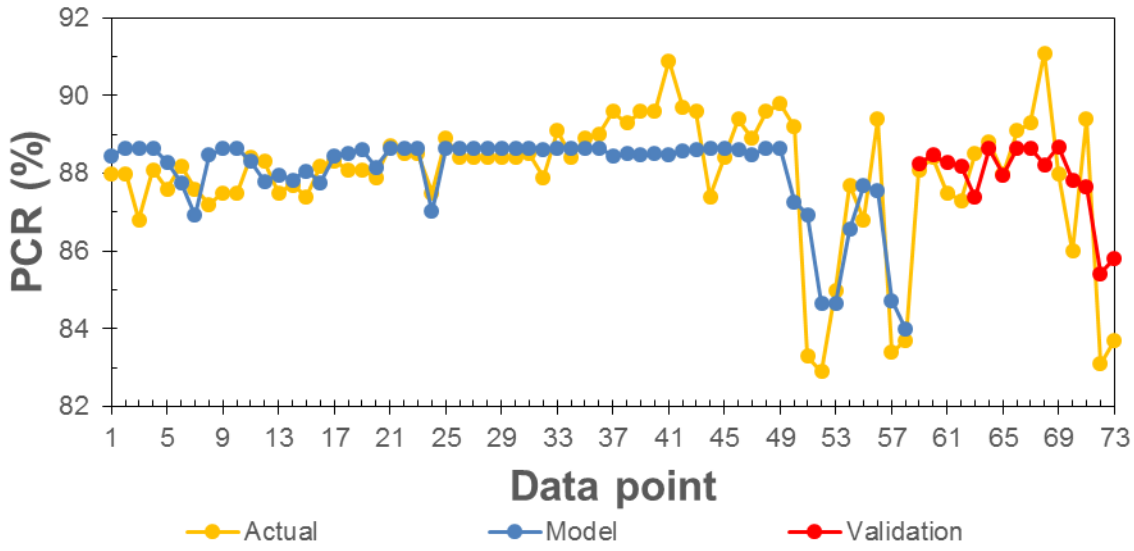


Figure 4-38: Phase 2 ANN-I PCR model performance

The PCR ANN-I model shows undertraining by predicting almost the same value for data points 25 to 49. Undertraining occurred due to early training stopping. The early training stop was due to the extreme high and low values in the validation dataset that resulted in the calculation of higher validation errors during the training phase. The higher ANN-I regression values are misleading due to some data points at the far left end of the PCR data range as shown in Figure 4-39. Most of the data points are situated in a horizontal fashion; however, the regression fit is adjusted due to the data points at the lower end of the dataset. The MLR-I model performed better compared to the ANN-I model, however both models are inaccurate for prediction purposes.

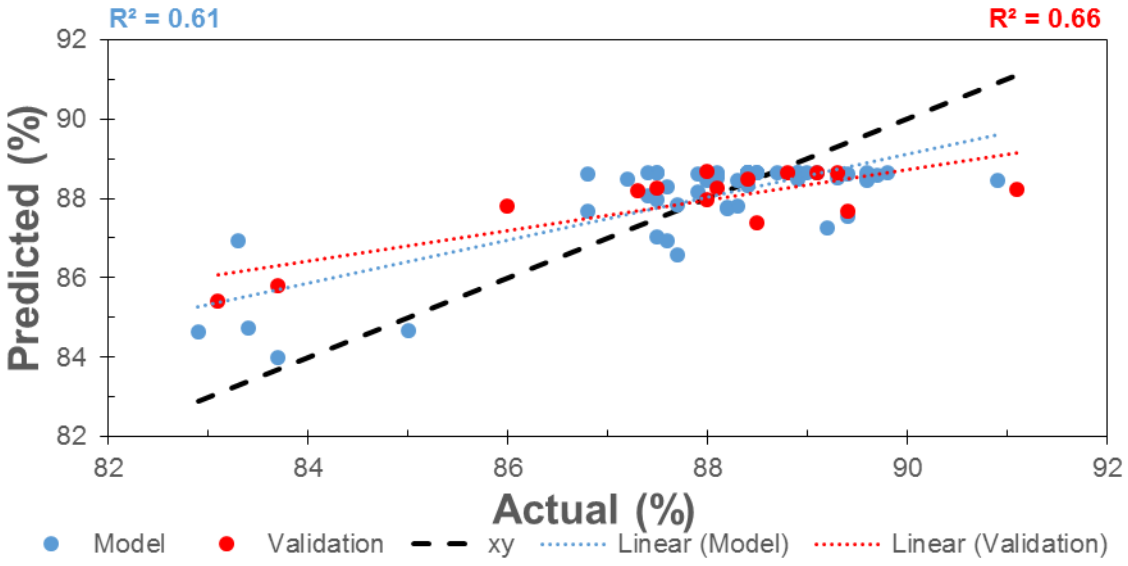


Figure 4-39: Phase 2 ANN-I PCR model regression

The ANN-I product particle slope (PSL) model has a higher model and validation regression fit with lower error values compared to the MLR-I model. This is depicted in Figure 4-40 and Figure 4-41. Both models have a moderate MAPE value due to large errors on data points 3, 6 and 32. The MLR-I model follows the general trend with a low accuracy. The model struggles to fit the high PSL values in the dataset, especially at data points 13, 41 and 68. The ANN-I model performed better on data points 13, 41 and 68. The ANN-I model follows the trend of the data with a higher accuracy but lack to fit the first half of the validation dataset. Both models exhibit average performances with low accuracies, resulting in models that are not completely fit for prediction.

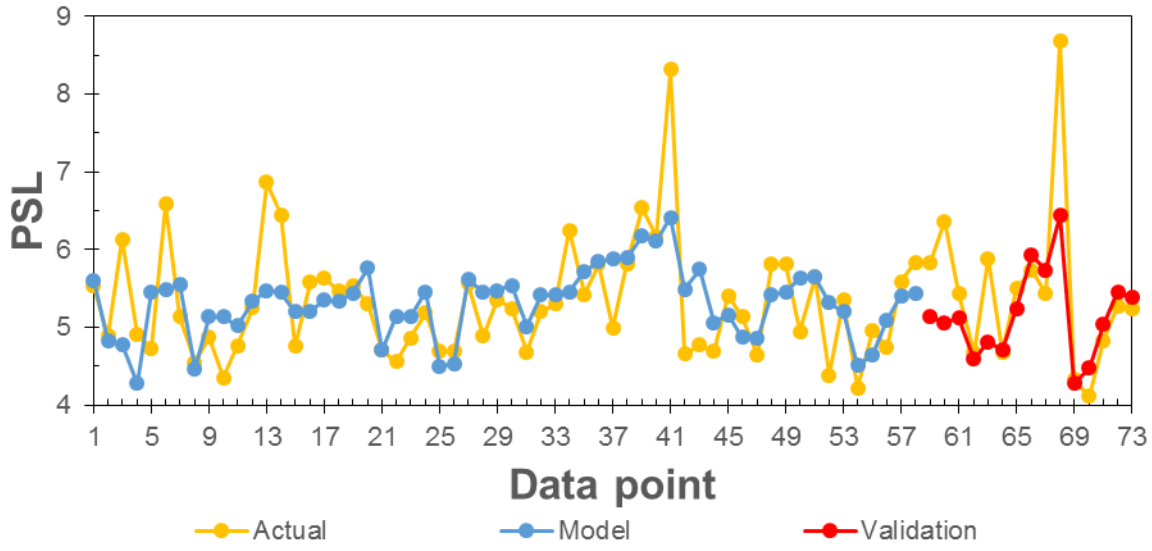


Figure 4-40: Phase 2 MLR-I PSL model performance

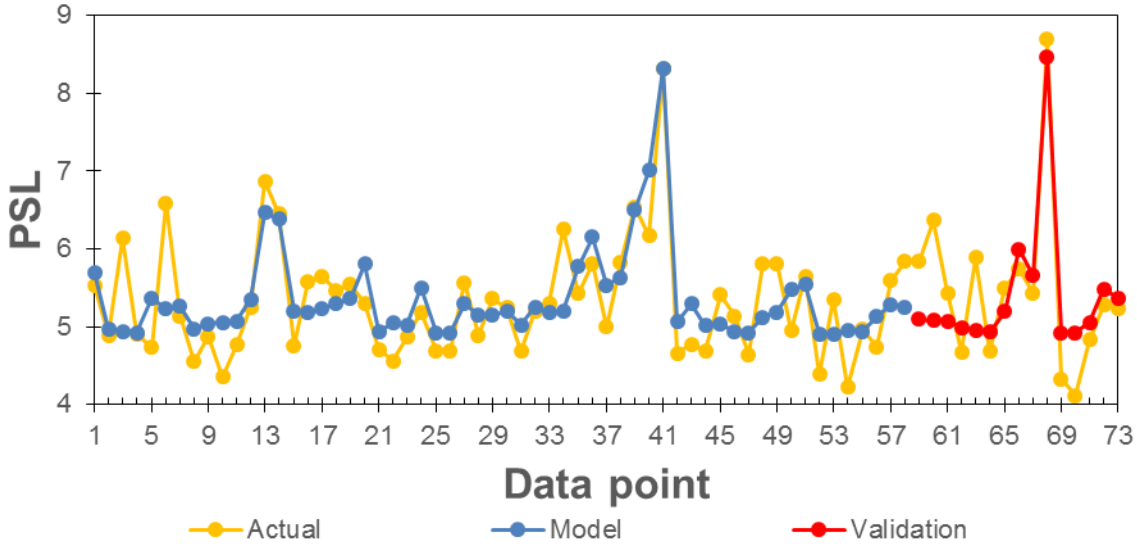


Figure 4-41: Phase 2 ANN-I PSL model performance

The MLR-I granulator particle slope (GSL) model performed similar to the ANN-I model. The ANN-I model has higher CMD values with lower error values compared to the MLR-I model. Both models follow the trend of the data with a high accuracy as shown in Figure 4-42 and Figure 4-43. The models also find it hard to fit the high GSL values at points 41 and 53 and some lower GSL values at points 26, 47 and 64. The ANN-I model fits the high GSL values with a higher accuracy but still struggles with the lower values. The ANN-I model performed better compared to the MLR-I model. Both models performed well with a fair prediction capability.

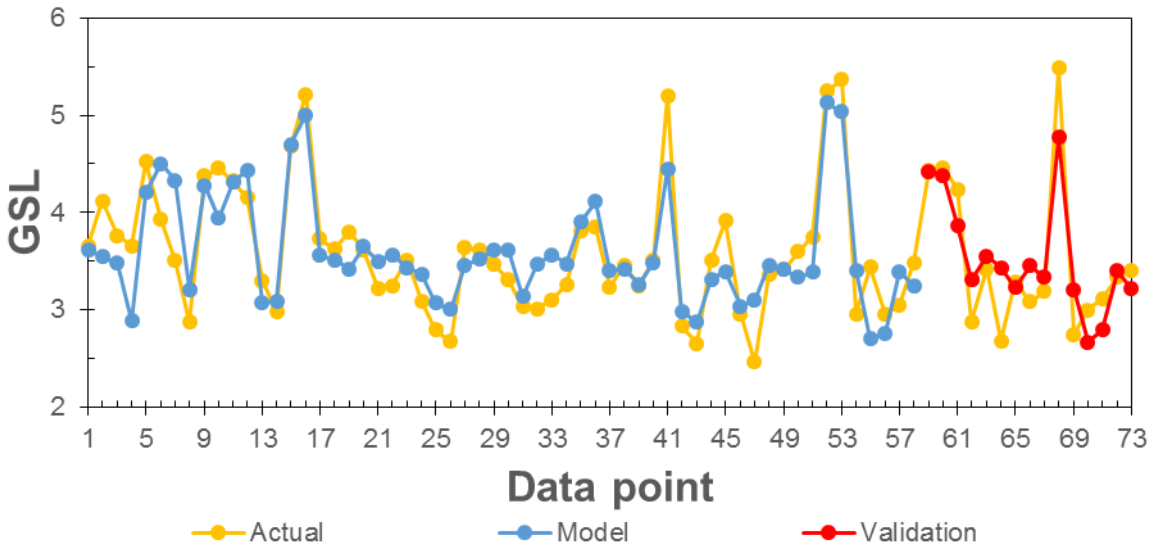


Figure 4-42: Phase 2 MLR-I GSL model performance

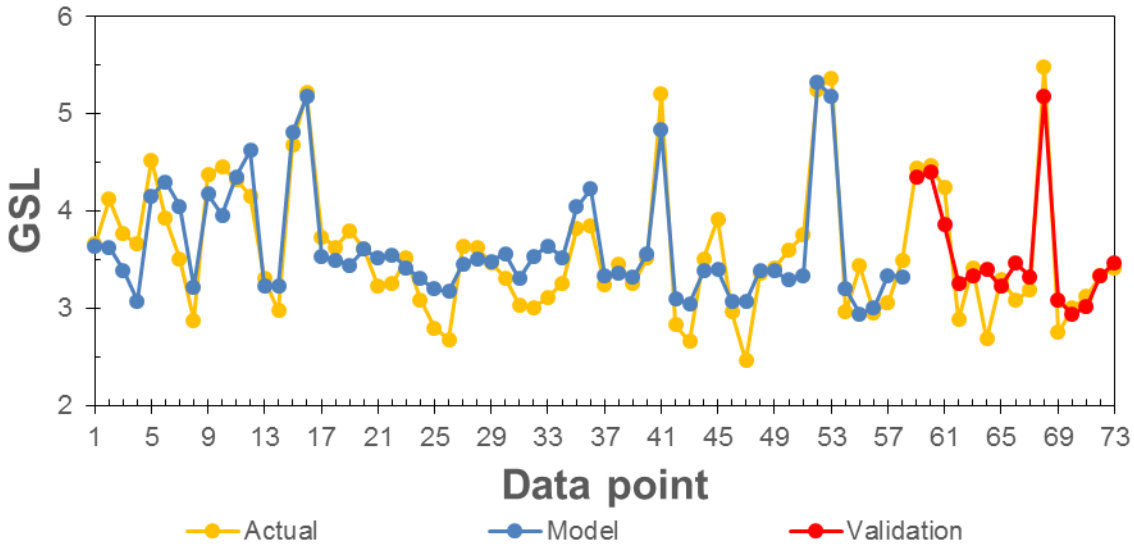


Figure 4-43: Phase 2 ANN-I GSL model performance

Both techniques developed accurate models with good prediction capabilities, especially the RUR, RER, EFF and GSL models. Undertraining occurred in most of the ANN-I models, limiting the performance of the models resulting in better performances from the MLR-I. A few of the ANN-I models still performed with a fair accuracy despite the undertraining.

The Spearman results indicated that it might be difficult to develop models for the GPS and POR models due to the low correlations between the operating and output variables. Some strong correlations were obtained for the PPS, PCR and PSL variables, but accurate models could not be developed.

The MLR-I modelling technique performed overall better than the unoptimised ANN-I models. Most MLR-I models had a good performance following the trend of the data with a fair accuracy. The ANN-I models were not optimised, resulting in lower performances. Optimisation of the ANN-I models will increase the accuracies and performances of the models.

Chapter 5. Conclusions and recommendations

5.1 Conclusions

Fluidised bed granulation is a non-linear process with various operating variables influencing the process and product properties, making it important to determine the significant variables. The Spearman's rho matrix is useful to determine the correlations between the different variables and to indicate the possible statistical significant independent variables for each dependent variable. Most of the correlations obtained correlates with literature observations.

It is concluded from the Spearman results that all the operating variables are independent from one another. The spray variables have the largest correlations with the output variables, making them important variables to control. The Spearman's correlation method proved successful as a preliminary step to determine the independent variables for model development. A larger dataset with more experimental levels improved the significance of the correlation values and showed good agreements with observations found in literature.

High accuracy models are developed using the both the main effects (MLR-M) and interaction effects (MLR-I) multiple linear regression (MLR) modelling technique. The granulator particle mean size (GPS) and product porosity (POR) dependent variables could not be modelled by just using the main effects, resulting in the inclusion of two-way interactions. It is concluded that the addition of two-way interactions did not influenced the significance of the MLR-I models, but increased the models' accuracies, indicating the importance of including the two-way interactions. Some of the main effects and interaction effects MLR models only followed the data trend with a low accuracy, making them not suitable for model prediction.

The MLR-I models performed better than some of the artificial neural network (ANN) models, developed using all the variables (ANN-I), due to insufficient training as a result of the early validation-check stop. It is concluded that these ANN-I models cannot follow the data trend and resulted in predicting almost constant values for certain data points. Some of the ANN-I models have a high performance with a sufficient accuracy, making them ideal for prediction. Good performing MLR-I and ANN-I models are obtained for the production rate (RUR), recycle ratio (RER), efficiency (EFF) and granulator particle slope (GSL) variables with high prediction capabilities at accuracies in excess of 70%. Accurate models could not be developed for the PPS, GPS, product circularity (PCR), POR and product particle slope (PSL) variables.

The main effects MLR technique is useful as a preliminary model development method to determine the statistical significant independent variables for the ANN modelling technique. The ANN-M models performed well with moderately high accuracies and low error values compared to ANN-I models as discussed in Appendix C. The size and complexity of the ANN-M models are reduced by using the main effects MLR technique. Including all the independent variables increased the number of weights, resulting in some ANN-I models experiencing undertraining.

5.2 Recommendations

The output variables, especially the RUR, RER and EFF variables should not be calculated using the operating variables. This gives false correlations due to the calculation methods, which skews the results. The output variables should be measured using independent measuring instruments. The results will be more accurate using weight values obtained from the respective conveyor belts from the final product and recycled product streams. The EFF variable should not be used as a quality parameter, as it cannot be directly measured.

The POR measuring technique is not accurate due to the oil solution adhering to the outside of the granules and loss of material during sample removal. The pore sizes and volumes can be determine more accurately using the Brunauer Emmett Teller (BET) technique. A gas pycnometer measuring technique is an alternative technique that can be used to determine the internal density of the granules. The internal density can then be used to determine the porosity of the granules. Both the BET and gas pycnometer methods are accurate and repeatable measuring techniques.

Determining the important operating variables is difficult to do, but it is crucial for accurate model development. The Spearman's matrix is a useful statistical technique to determine the variables that have the largest impact for a given dependent variable. Using only the important operating variables will improve the accuracy and significance of the models and will reduce the size of the datasets required. Other operating variables that can also be investigated include the bed height, bed density and granulator humidity.

A particle growth efficiency assessment can be investigated by crash stopping the granulator at certain time intervals and taking samples axially throughout the granulator. These samples will give a representative model of the granule growth in the granulator.

Future work can include the development of a control system that uses the developed models as inference measurements. The dynamic performance of the models along with the performance of the control system can be investigated and evaluated.

The ANN modelling technique was only investigated qualitatively leading to the use of unoptimised ANN models. Future work can include investigating whether the performance and accuracies can increased by optimising the ANN models and using larger model and validation datasets. The final models can be used to develop an inference control system. The performance of this system can be investigated and evaluated.

Visual observations indicated low fluidisation velocities in the granulator bed. This resulted in the formation of dead zones, especially at the sidewalls of the granulator. Using fans that can supply higher flow rates will increase the fluidisation of the granulator bed, thereby improving the mixing and granulation. Redesigning the perforated plate or granulator walls will also decrease the formation of dead zones and improve the fluidisation of the granulator bed.

Nozzle blocking and lump formation was a recurring problem due to low fluidisations and dead zones blocking the spray path, resulting in wet quenching and lumping to occur. Determining the nozzle spray pattern and fluidisation pattern will help to identify the correct type of nozzle to be used. Improved fluidisation and operation with the efficient spray nozzle will increase the production rate and lower the risk of lump formation.

The sampling port near the feed to the granulator samples the product inside the granulator. These samples are not a true representation of the seed material due to the fluidisation and recirculation of material inside the bed. The sampling port underneath the crusher only represents the oversize particles that are crushed. Using this sampling port is a better representative of the seed material; however, the undersize recycling stream is omitted. A sampling port containing both streams is required to accurately measure the seed particle material entering the granulator.

Chapter 6. References

Abraham, A., 2004. Meta learning evolutionary artificial neural networks. *Neurocomputing*, 56, pp.1–38.

Aleksic, I. et al., 2015. An investigation into the usefulness of different empirical modeling techniques for better control of spray-on fluidized bed melt granulation. *International journal of pharmaceutics*, 496(2), pp.627–635.

Aleksic, I. et al., 2014. In silico modeling of in situ fluidized bed melt granulation. *International journal of pharmaceutics*, 466(1), pp.21–30.

Ayat, S. & Pour, Z.A., 2014. Comparison between artificial neural network learning algorithms for prediction of student average considering effective factors in Learning and educational progress. *Journal of mathematics and computer science*, 8, pp.215–225.

Azadi, S. & Karimi-Jashni, A., 2016. Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of Fars province, Iran. *Waste Management*, 48, pp.14–23.

Beale, M., Hagan, M.T. & Demuth, H.B., 1992. *Neural network toolbox*,

Beale, M.H., Hagan, M.T. & Demuth, H.B., 2014. *Neural network toolbox user's guide R2014a*, The Math Works Inc.

Becher, R.-D. & Schlünder, E.-U., 1998. Fluidized bed granulation—the importance of a drying zone for the particle growth mechanism. *Chemical Engineering and Processing: Process Intensification*, 37(1), pp.1–6.

Berenson, M. et al., 2012. *Basic business statistics: Concepts and applications* 12th ed., Pearson Higher Education AU.

Berenson, M. & Levine, D., 1989. *Basic business statistics: Concepts and applications* 4th ed., Pearson Higher Education AU.

Biswal, S.S., 2011. Particle Growth and Film coating in a Fluidized Bed Granulator. *Report*.

Burggraeve, A. et al., 2012. Development of a fluid bed granulation process control strategy based on real-time process and product measurements. *Talanta*, 100, pp.293–302.

Burggraeve, A. et al., 2013. Process analytical tools for monitoring, understanding, and control of pharmaceutical fluidized bed granulation: A review. *European journal of pharmaceutics and biopharmaceutics*, 83(1), pp.2–15.

Burggraeve, A. et al., 2011. Understanding fluidized-bed granulation. *Pharmaceutical Technology*, 35(8), pp.63–67.

Cameron, I. et al., 2005. Process systems modelling and applications in granulation: A review. *Chemical Engineering Science*, 60(14), pp.3723–3750.

Cohen, J., 1992. A power primer. *Psychological bulletin*, 112(1), p.155.

Devore, J. & Farnum, N., 2005. *Applied statistics for engineers and scientists* 2nd ed., Brooks and Cole.

Ellis, S. & Steyn, H., 2003. Practical significance (effect sizes) versus or in combination with statistical significance (p-values): Research note. *Management Dynamics: Journal of the Southern African Institute for Management Scientists*, 12(4), p.p–51.

Field, A., 2009. *Discovering statistics using SPSS* 3rd ed., Sage publications.

Fries, L. et al., 2014. Product design based on discrete particle modeling of a fluidized bed granulator. *Particuology*, 12, pp.13–24.

Ghaffari, A. et al., 2006. Performance comparison of neural network training algorithms in modeling of bimodal drug delivery. *International journal of pharmaceutics*, 327(1), pp.126–138.

Hauke, J. & Kossowski, T., 2011. Comparison of values of Pearson's and Spearman's correlation coefficient on the same sets of data. *Quaestiones Geographicae*, 30(2).

Heinrich, S. et al., 2003. Particle population modeling in fluidized bed-spray granulation—analysis of the steady state and unsteady behavior. *Powder technology*, 130(1), pp.154–161.

Hill, T. & Lewicki, P., 2006. *Statistics: methods and applications: a comprehensive reference for science, industry, and data mining* 1st ed., StatSoft, Inc.

Hinkelmann Klaus, O.K., 1994. *Design and Analysis of Experiments*, John Wiley & Sons.

Hyndman, R.J., 2014. Measuring forecast accuracy.

Hyndman, R.J. & Koehler, A.B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), pp.679–688.

Iveson, S.M. et al., 2001. Nucleation, growth and breakage phenomena in agitated wet granulation processes: a review. *Powder technology*, 117(1), pp.3–39.

Jain, A.K., Mao, J. & Mohiuddin, K.M., 1996. Artificial neural networks: A tutorial. *IEEE computer*, 29(3), pp.31–44.

Jayaraman, K., 2000. *A statistical manual for forestry research*, Forestry Research Support Programme for Asia and the Pacific.

Karsoliya, S., 2012. Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. *International Journal of Engineering Trends and Technology*, 3(6),

pp.713–717.

Korteby, Y. et al., 2016. Implementation of an artificial neural network as a PAT tool for the prediction of temperature distribution within a pharmaceutical fluidized bed granulator. *European Journal of Pharmaceutical Sciences*, 88, pp.219–232.

Kriesel, D., 2007. *A brief Introduction on Neural Networks*, Available at: http://www.dkriesel.com.

Landau, S. & Everitt, B., 2004. *A handbook of statistical analyses using SPSS* 1st ed., Chapman & Hall/CRC.

Legendre, P. & Legendre, L., 1998. *Numerical ecology: second English edition* 2nd ed., Elsevier.

Macias-Garcia, A., Cuerda-Correa, E.M. & Diaz-Diez, M., 2004. Application of the Rosin-Rammler and Gates-Gaudin-Schuhmann models to the particle size distribution analysis of agglomerated cork. *Materials Characterization*, 52(2), pp.159–164.

Maria, S., 2010. Methods for porosity measurement in lime-based mortars. *Construction and Building Materials*, 24(12), pp.2572–2578.

Montgomery, D.C., 2008. *Design and analysis of experiments* 8th ed., John Wiley & Sons.

Moraga, S.V. et al., 2015. Fluidized-bed melt granulation: The effect of operating variables on process performance and granule properties. *Powder Technology*, 286, pp.654–667.

Murtoniemi, E. et al., 1994. The advantages by the use of neural networks in modelling the fluidized bed granulation process. *International journal of pharmaceutics*, 108(2), pp.155–164.

Napier-Munn, T. & Wills, B.A., 2006. *Wills' mineral processing technology: An introduction to the practical aspects of ore treatment and mineral recovery* 7th ed., Elsevier.

Oehlert, G.W., 2010. *A first course in design and analysis of experiments*, University of Minnesota.

Palis, S., Bück, A. & Kienle, A., 2012. Low order controller design for continuous fluidized bed spray granulatiovn with internal product classification by robust control methods. *Proceedings of the 7th IFAC symposium on Robust Control Design*, 45(13), pp.701–706.

Pashminehazar, R., Kharaghani, A. & Tsotsas, E., 2016. Three dimensional characterization of morphology and internal structure of soft material agglomerates produced in spray fluidized bed by X-ray tomography. *Powder Technology*.

Patel, T.B. et al., 2010. Influence of process variables on physicochemical properties of the granulation mechanism of diclofenac sodium in fluid bed granulation. *International Journal of Pharmaceutical Sciences Review and Research*, 3(1), pp.61–65.

Qiu, Y. et al., 2009. *Developing solid oral dosage forms: pharmaceutical theory & practice* 1st ed., Academic press.

Rambali, B., Baert, L. & Massart, D.L., 2001. Using experimental design to optimize the process parameters in fluidized bed granulation on a semi-full scale. *International journal of pharmaceutics*, 220(1-2), pp.149–60.

Rhodes, M.J., 2008. *Introduction to particle technology* 2nd ed., John Wiley & Sons.

Sahoo, M.L., 2012. *Studies on Effect of Process Parameters on Particle Growth in a Fluidized Bed Granulator*.

Satish, S. & Setty, Y.P., 2005. Modeling of a continuous fluidized bed dryer using artificial neural networks. *International Communications in Heat and Mass Transfer*, 32(3), pp.539–547.

Seborg, D.E. et al., 2011. *Process Dynamcs and Control* 3rd ed., John Wiley & Sons.

Sharma, P. et al., 2013. Feedforward neural network: A review. *International Journal of Advanced Research in Engineering and Applied Sciences*, 2(10).

Singh, V., Dubey, M. & Singh, A., 2011. Study the Effect of Process Variables in Fluid Bed Granulation on the Physico-Chemical Properties of Granules. *IJPSR*, 2(2), pp.370–377.

Smith, P. & Nienow, A., 1983. Particle growth mechanisms in fluidised bed granulation—I: the effect of process variables. *Chemical Engineering Science*, 38(8), pp.1223–1231.

Srinivasakannan, C. & Balasubramaniam, N., 2003. Particle growth in fluidised bed granulation. *Chemical and biochemical engineering quarterly*, 17(3), pp.201–206.

Srivastava, S. & Mishra, G., 2010. Fluid bed technology: overview and parameters for process selection. *International Journal of Pharmaceutical Sciences and Drug Research*, 2(4), pp.236–246.

Topuz, A., 2010. Predicting moisture content of agricultural products using artificial neural networks. *Advances in Engineering Software*, 41(3), pp.464–470.

Vivaracho-Pascual, C. et al., 2016. Client threshold prediction in biometric signature recognition by means of Multiple Linear Regression and its use for score normalization. *Pattern Recognition*, 55, pp.1–13.

Webb, P.A., 2001. Volume and density determinations for particle technologists. *Micromeritics Instrument Corp*, 2(16), p.01.

Wilamowski, B.M. & Irwin, J.D., 2011. *Industrial electronics handbook - Intelligent systems* 2nd ed., CRC Press.

Wilamowski, B.M. & Yu, H., 2010. Neural network learning without backpropagation. *IEEE Transactions on Neural Networks*, 21(11), pp.1793–1803.

Willmott, C.J. & Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), pp.79–82.

Wong, P.M., Chan, L.W. & Heng, P.W.S., 2013. Investigation on side-spray fluidized bed granulation with swirling airflow. *AAPS PharmSciTech*, 14(1), pp.211–221.

Yao, Y. et al., 2009. Non-destructive characterization of coal samples from China using microfocus X-ray computed tomography. *International Journal of Coal Geology*, 80(2), pp.113–123.

Zhang, J. et al., 2000. Evaluation of control strategies for fertiliser granulation circuits using dynamic simulation. *Powder Technology*, 108(2), pp.122–129.

Zhu, Z. et al., 2011. Local linear regression for soft-sensor design with application to an industrial deethanizer. *IFAC Proceedings Volumes*, 44(1), pp.2839–2844.

Ziyani, L. & Fatah, N., 2014. Use of experimental designs to optimize fluidized bed granulation of maltodextrin. *Advanced Powder Technology*, 25(3), pp.1069–1075.

Appendix A - Calculations

Roslin-Ramler linearization

The slope (*m*) was calculated by the linearization of the Roslin-Ramler equation (Macias-Garcia et al. 2004). The Roslin-Ramler equation is given by Equation A-1 where the natural logarithmic is taken on both sides twice to give Equation A-2 and Equation A-3. *P(d)* represents the distribution function with d_i representing the particle size for the i[th] percentile.

$$P(d) = 1 - \exp\left[-\left(\frac{d}{d_{63.2}}\right)^{m}\right] \quad \textbf{Equation A-1}$$

$$\ln[1 - P(d)] = -\left(\frac{d}{d_{63.2}}\right)^{m} \quad \textbf{Equation A-2}$$

$$\ln\{-\ln[1 - P(d)]\} = m.\ln(d) - m.\ln(d_{63.2}) \quad \textbf{Equation A-3}$$

The Roslin-Ramler equation in now linearized where the $\ln\{-\ln[1 - P(d)]\}$ term is the y-term, $\ln(d)$ the x-term and $-m.\ln(d_{63.2})$ the intercept with the y-axis. The slope can be calculated using two points according to Equation A-4 and Equation A-5.

$$m = \frac{\ln\{-\ln[1 - P(d_2)]\} - \ln\{-\ln[1 - P(d_1)]\}}{\ln(d_2) - \ln(d_1)} \quad \textbf{Equation A-4}$$

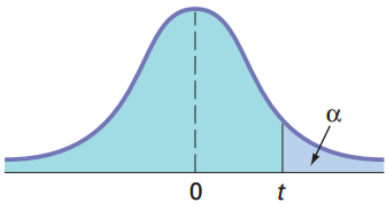
$$m = \ln\left\{\frac{\ln[1 - P(d_{25})]}{\ln[1 - P(d_{75})]}\right\} \div \ln\left(\frac{d_{25}}{d_{75}}\right) \quad \textbf{Equation A-5}$$

Repeatability

A 95% confidence interval was chosen to calculate the experimental error since all the variables in the models are also tested with a 95% confidence. Five repeat runs of each phase were conducted resulting in a t constant (t_{n-1}) value of 2.776, obtained from Figure A-1. The error and average percentage error are calculated using Equation A-6 and Equation A-7 respectively where *n* represents the sample size (Devore & Farnum 2005, pp.313–319).

$$Error = t_{n-1}\left(\frac{Standard\ deviation}{\sqrt{n}}\right)$$ **Equation A-6**

$$Average\ percentage\ error = \frac{Error}{\bar{Y}}$$ **Equation A-7**



Degrees of Freedom	Cumulative Probabilities					
	0.75	0.90	0.95	0.975	0.99	0.995
	Upper-Tail Areas					
	0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693

Figure A-1: t critical values for confidence intervals taken from Berenson et al. (2012, p.800)

ANN training algorithm

The ANN algorithm starts with a feed forward process where the net input (net_{oj}) for each output (*o*) for each layer (*j*) is calculated using Equation A-8, where W_{ioj} represents the weight value for input neuron *i* to output neuron *o* for layer *j*. The bias term always has a value of 1. The node value is calculated by using an activation function (*f*) on the net input value using Equation A-9. The activation function in Equation A-9 is the sigmoid function for all the layers except the last layer where a linear function is used. An ANN without bias nodes are shown in Figure A-2 where some connection and notations are left out due to illegibility. The error (*e*) can be calculated using Equation A-10 after the final output neurons are calculated (Abraham 2004; Wilamowski & Irwin 2011, p.7; Wilamowski & Yu 2010).

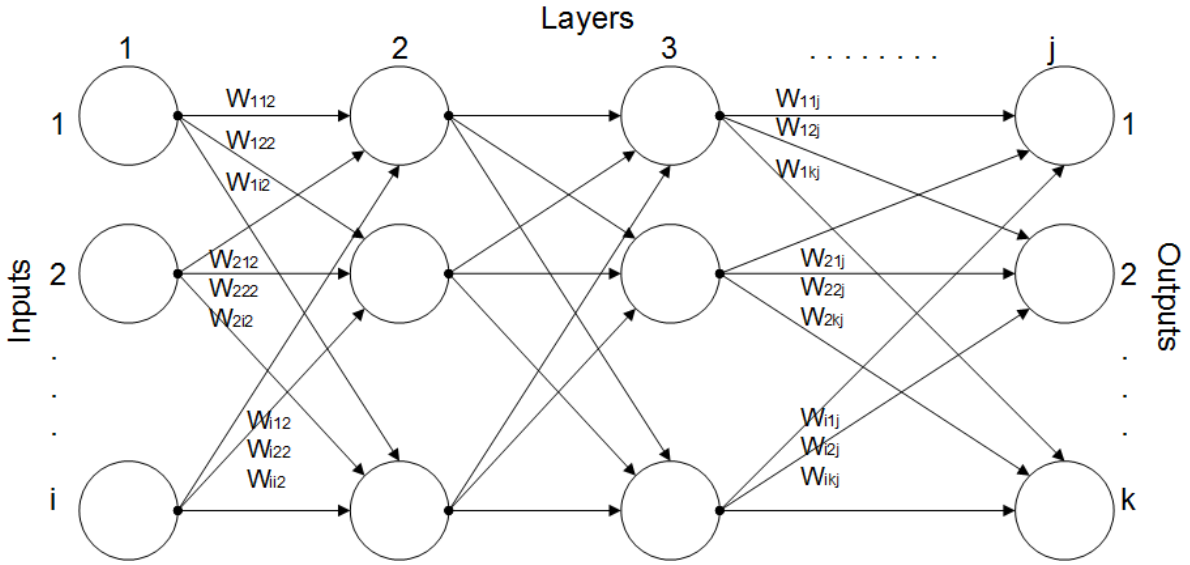


Figure A-2: An illustration of an artificial neural network topography adapted from Wilamowski & Irwin (2011, p.10).

$$net_{oj} = \sum_{i=1}^{m} W_{ioj} X_i + bias_j$$ **Equation A-8**

$$Node\ value = f\left(net_{oj}\right)$$ **Equation A-9**

$$e_j = Desired_j - Output_j$$ **Equation A-10**

The Levenberg-Marquardt training algorithm starts by calculating the slopes of each layer using Equation A-11. The derivative of the error is calculated using Equation A-12 forming the elements for the Jacobian matrix. The weights are then updated using Equation A-13 where the Jacobian matrix and the previous weight matrix (***W***) with size *N* x 1 are used. *N* represents the total number of weights. The Jacobian matrix (***J***) is a (*n* x *k*) x *N* matrix, where

n represents the training samples size and *k* the total amount of independent, output, variables. The identity matrix (***I***) is an *N* x *N* matrix and the error matrix (***e***) is an (*n* x *k*) x 1 matrix ((Wilamowski & Irwin 2011, p.7; Wilamowski & Yu 2010).

$$slope_j = \frac{\partial Output_j}{\partial net_j}$$ **Equation A-11**

$$\frac{\partial e_{m,j}}{\partial w_{m,j,i}} = -f'(net_j) \times s_j \times Output_j$$ **Equation A-12**

$$\boldsymbol{W}_{K+1} = \boldsymbol{W}_K - (\boldsymbol{J}_K^T \boldsymbol{J}_K + \mu \boldsymbol{I})^{-1} \boldsymbol{J}_K^T \boldsymbol{e}_K$$ **Equation A-13**

Appendix B - Experimental Planning

Phase one experimental planning

Table B-1 contains the abbreviations, description and units for each operating variable. The operating variable ranges and level values are given in Table B-2. Only 2 levels of the spray liquid concentration (SLC) variable could be obtained, 90% and 92%. The experimental plan for the first phase can be find in Table B-3.

Table B-1: Abbreviations of process variables

Abbreviation	Description	Unit
FAF	Fluidised air flow rate	Nm^3/h
FAT	Fluidised air temperature	°C
SLF	Spray liquid flow rate	tph
SLT	Spray liquid temperature	°C
SLC	Spray liquid concentration	%
SPS	Seed particle size d_{50}	mm

Table B-2: Phase one variable ranges

	FAF	FAT	SLF	SLT	SLC	SPS
Min value	9 000	90	4	120	90	1
Max value	13 000	135	9	133	92	1.4

Table B-3: Phase one experimental plan

Run	FAF	FAT	SLF	SLT	SLC	FSD
1	13000	112.5	9	126.5	92	1
2	13000	90	9	133	92	1.2
3	11000	112.5	6.5	126.5	90	1.2
4	13000	112.5	6.5	120	90	1.4
5	9000	135	6.5	126.5	90	1
6	13000	112.5	6.5	126.5	92	1
7	9000	135	4	133	92	1.4
8	11000	90	9	120	92	1.2
9	13000	112.5	6.5	126.5	92	1

Phase two experimental planning

More levels were investigated in the second phase. Due to plant limitations, only three concentration levels were investigated, 90%, 92% and 94%. Table B-4 summarises the ranges for each operating condition.

Table B-4: Phase two variable ranges

	FAF	FAT	SLF	SLT	SLC	SPS
Min value	9 000	68	4	121	90	1
Max value	13 000	135	9	151	94	1.4

Nozzle blocking was a problem from phase one due to inappropriate spray liquid temperatures (SLTs) for the different concentration levels. A fixed SLT range for all the concentrations could not be obtained. The only solution was to use a different SLT range for each concentration level and creating five levels for the SLT from each range. Table B-5 summarises the SLT values for each concentration level.

Table B-5: Phase two spray liquid temperatures for different concentrations

		SLT						
		Min range	Max range	Level 1	Level 2	Level 3	Level 4	Level 5
SLC	90%	121	134	121	124	128	131	134
	92%	127	141	127	131	134	138	141
	94%	136	151	136	140	144	147	151

The experimental plan was developed as discussed in section 3.2.2 where the levels were assigned randomly to each variable in order to produce experimental data with varying operating variable changes from each run. Table B-6 summarises the experimental plan.

Table B-6: Phase two experimental plan

Run	FAF	FAT	SLF	SLT	SLC	SPS
1	11 000	135	8	136	94	1.2
2	12 000	124	4.6	151	94	1.3
3	13 000	113	3.5	140	94	1.3
4	12 000	113	8	140	94	1.1
5	12 000	124	4.6	151	94	1.3
6	10 000	101	6.9	144	94	1
7	12 000	124	4.6	151	94	1.3
8	9 000	124	4.6	144	94	1.2
9	13 000	90	8	147	94	1.4
10	11 000	135	4.6	147	94	1.4
11	12 000	124	4.6	151	94	1.3
12	13 000	90	3.5	140	94	1.1
13	12 000	124	4.6	151	94	1.3
14	11 000	101	8	151	94	1.2
15	12 000	101	6.9	141	92	1.2
16	11 000	101	4.6	134	92	1.3
17	11 000	113	8	138	92	1.4
18	12 000	113	4.6	134	92	1.3
19	9 000	124	5.8	138	92	1
20	10 000	90	5.8	141	92	1.2
21	13 000	124	8	131	92	1.2
22	10 000	135	3.5	127	92	1.1
23	10 000	135	6.9	131	92	1.4
24	12 000	101	8	134	90	1.1
25	10 000	124	6.9	131	90	1.3
26	11 000	135	6.9	134	90	1.1
27	9 000	113	4.6	131	90	1.3
28	9 000	124	5.8	128	90	1.2
29	11 000	101	5.8	128	90	1.2
30	10 000	135	6.9	121	90	1
31	12 000	90	5	124	90	1
32	11 000	68	8	151	94	1.2
33	9 000	79	8	140	94	1.3
34	12 000	113	6.9	136	94	1
35	12 000	90	8	127	92	1.1
36	12 000	68	6.9	131	92	1.4
37	10 000	113	4.6	131	92	1.3
38	11 000	90	3.5	134	92	1.4
39	10 000	79	5.8	141	92	1.2
40	13 000	68	4.6	134	90	1.4
41	11 000	135	6.9	134	90	1.1

Appendix C - Artificial neural network comparison

The neural networks were developed as discussed in section 3.3.3. The artificial neural network (ANN) models developed from the results from the main effects multiple linear regression (MLR) technique (ANN-M) were evaluated against ANN models developed using all the independent variables as inputs (ANN-I). The results for the neural network models are summarised in Appendix D.

Table C-1: Artificial neural network results for phase 2

	Parameter	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
Main effects (ANN-M)	Independent variables	2	2	2	3			2	3	5
	Model CMD	0.88	0.69	0.71	0.29			0.28	0.54	0.82
	Model ACD	0.88	0.68	0.70	0.25			0.26	0.51	0.81
	Validation CMD	0.91	0.87	0.65	0.51			0.05	0.77	0.90
	Overall CMD	0.89	0.72	0.69	0.30			0.21	0.61	0.84
	SEE	0.43	1.82	5.95	0.13			1.36	0.50	0.29
	MAE	0.29	1.26	4.44	0.10			0.90	0.41	0.22
	MAPE, %	9.6	13.8	6.9	5.1			1.0	7.7	6.4
	Best epochs	4	6	5	3			6	15	21
All variables (ANN-I)	Independent variables	7	7	7	7	7	7	7	7	7
	Model CMD	0.92	0.85	0.76	0.36	0.27	0.31	0.61	0.59	0.78
	Model ACD	0.91	0.82	0.72	0.27	0.17	0.21	0.55	0.54	0.75
	Validation CMD	0.94	0.80	0.51	0.55	0.23	0.20	0.66	0.72	0.89
	Overall CMD	0.93	0.83	0.67	0.36	0.26	0.29	0.61	0.64	0.80
	SEE	0.39	1.32	5.65	0.13	0.11	2.33	1.06	0.49	0.33
	MAE	0.28	0.98	4.08	0.10	0.08	1.79	0.75	0.37	0.26
	MAPE, %	8.6	11.8	6.1	4.9	5.1	4.9	0.9	6.9	7. 6
	Best epochs	17	5	4	8	16	6	19	15	13

The production rate (RUR) neural network model developed with all the variables has higher model, validation and overall coefficient of multiple determination (CMD) values. The model has a high overall accuracy with a 7% overall regression variation. The difference in the adjusted coefficient of multiple determination (ACD) values is low due to the high model regression values obtain. The standard error of estimate (SEE) and mean absolute error (MAE) values are also lower for the model with all the independent variables, resulting in a very good performance as seen in Figure C-2. Both models predict the values close to the actual data points and follow the trends well as shown in Figure C-1 and Figure C-2.

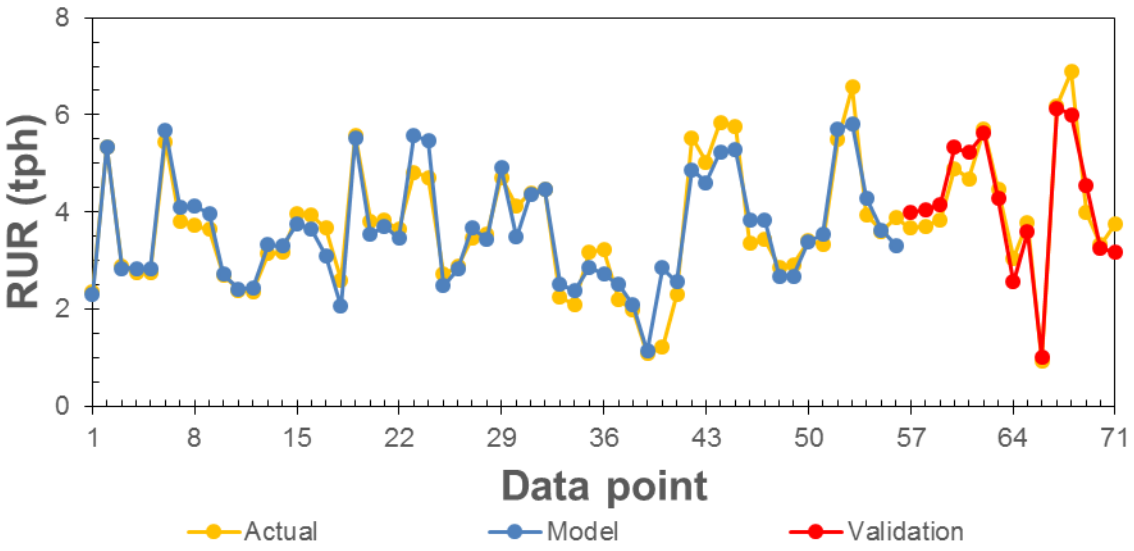


Figure C-1: Phase 2 ANN-M RUR model performance

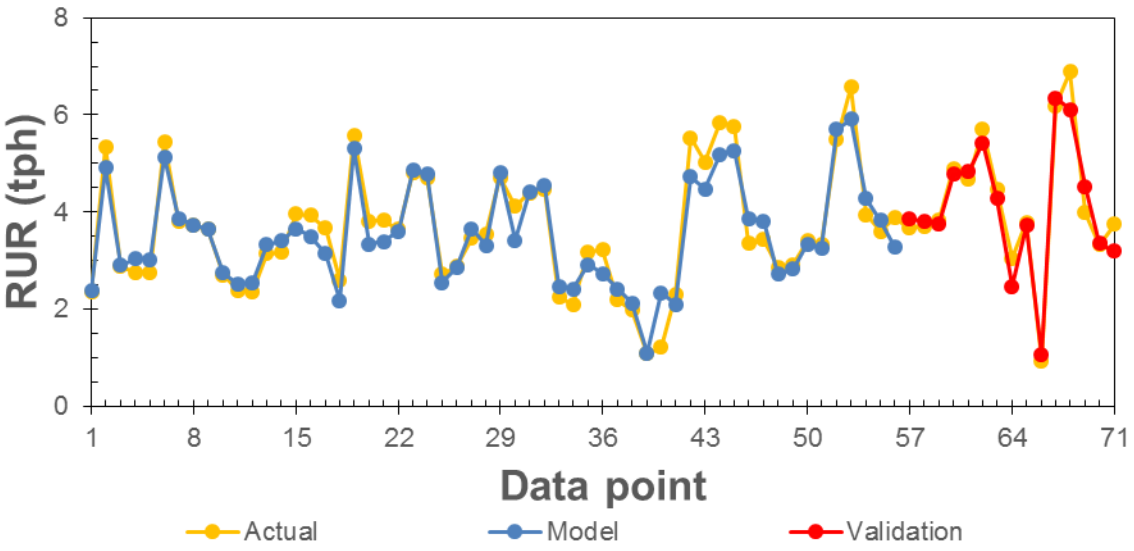


Figure C-2: Phase 2 ANN-I RUR model performance

Both RUR models have moderate mean absolute percentage error (MAPE) values due to some positive errors, e.g. data points 23 and 24 for the ANN-M model and data points 40 and 46 for the ANN-I model. The MAPE penalises positive errors resulting in larger MAPE values (Hyndman & Koehler 2006). The ANN-I model fits the validation data with a higher accuracy compared to the ANN-M model. The model also trained for 13 epochs longer as a result from the added variables. Both models follow the data trend well with a high accuracy, apart from a few data points, e.g. data points 42 to 45. The ANN- model fits the validation data almost precisely. Both ANN models performed well and are accurate enough to be used for prediction purposes.

The recycle ratio (RER) ANN model with only the main effects variables has a significantly lower regression fit on the modelling data, compared to the model with all the variables. The model has a higher fit on the validation data with a 7% increase validation CMD value. The ANN-I model has lower SEE and MAE values, making it more accurate. Both models have a moderately large MAPE value that is due to the large prediction errors on certain values, especially data points 36 and 44 as shown in Figure C-3 and Figure C-4.

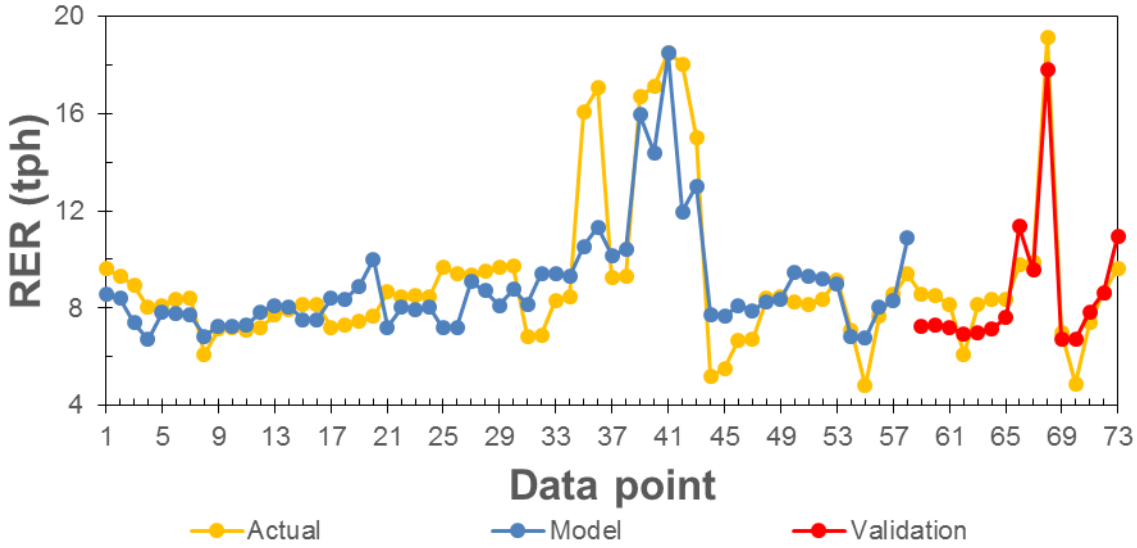


Figure C-3: Phase 2 ANN-M RER model performance

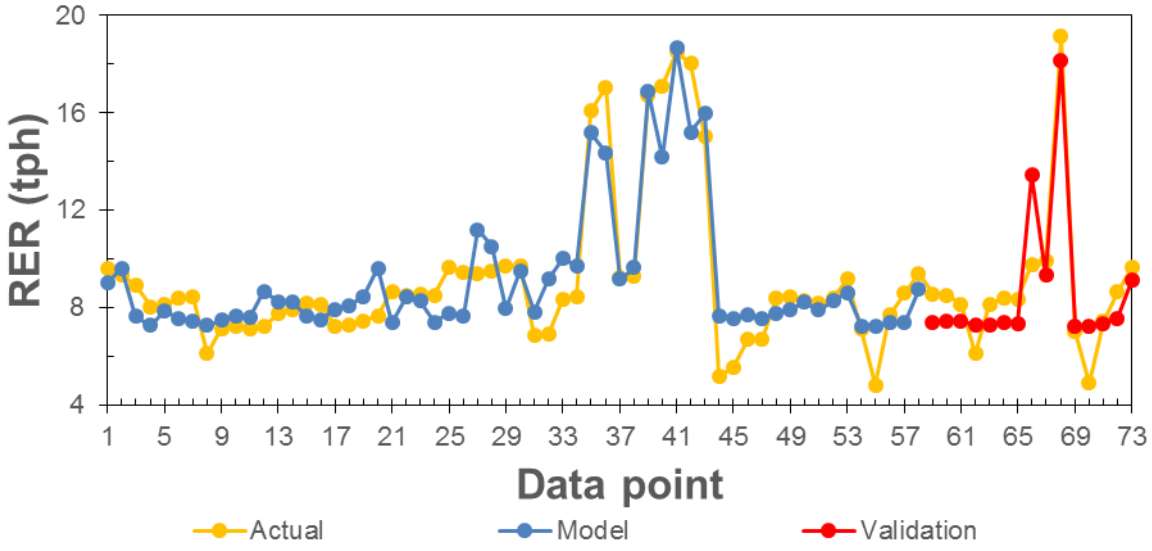


Figure C-4: Phase 2 ANN-I RER model performance

The ANN RER model with all the variables fits these points closer but still has a large calculated MAPE value due to the other high error values. Both models fit the average data points well but struggles with the extreme data points at 35 and 39 to 43. The ANN-I model follows the trend of the modelling data with a moderate accuracy, but struggles to follow the validation trend. The high data points at 35, 36, 39 to 43 and 68 makes it difficult for the models to accurately fit and predict these values. The validation dataset only contains one high data point, resulting in large errors being calculated during the validation of the training phase. This resulted in early stopping of the training phase as indicated by the near horizontal prediction at data points 44 to 50, 54 to 55 and 59 to 66. This near horizontal prediction results indicates undertraining; however, the models still followed the general trend with a fair accuracy. The regression values for the ANN-I model are in the same order, indicating that the model, validation and overall performances are in line with each other. This signifies the model's ability to produce predictions within the same level of accuracy. The ANN-I model is preferred due to its better performance and lower error values.

The efficiency (EFF) ANN model with all the variables has a higher model CMD but with a 10% lower validation coefficient compared to the main effects model. This model has the highest SEE and MAE values between all the other ANN models. The deviation between the actual and predicted values is seen in Figure C-5 and Figure C-6.

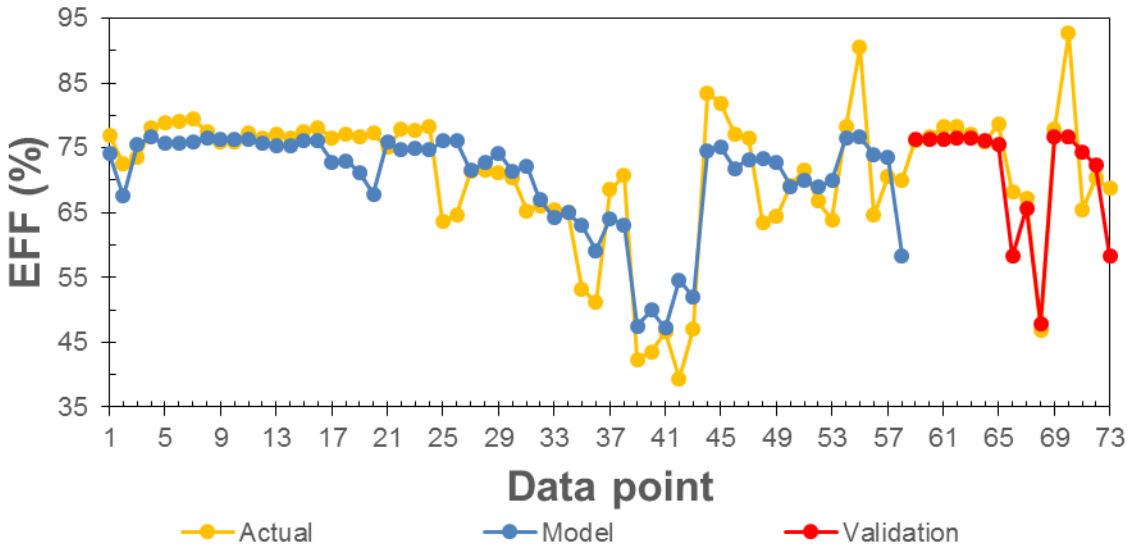


Figure C-5: Phase 2 ANN-M EFF model performance

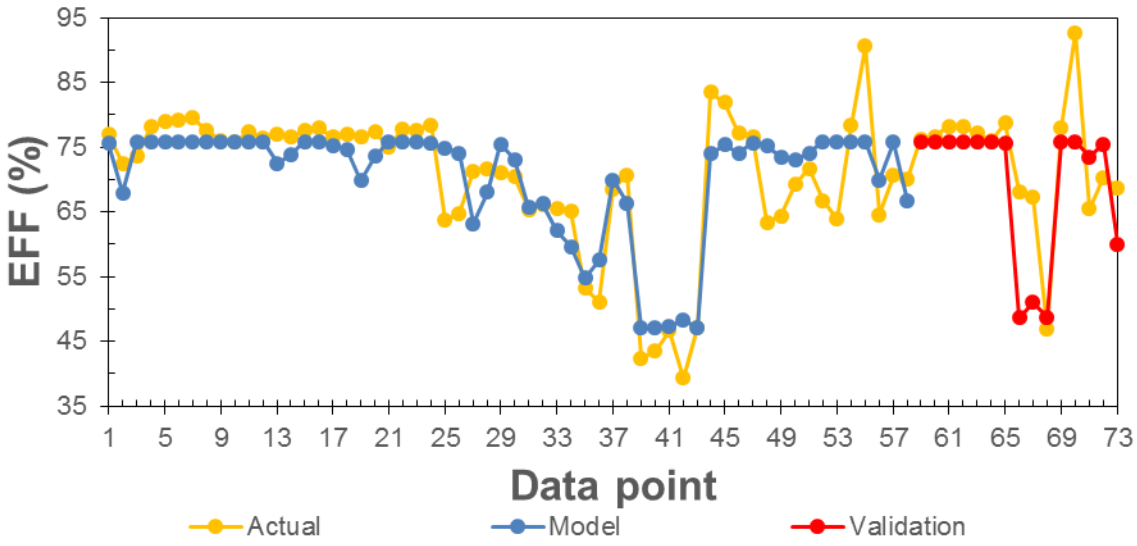


Figure C-6: Phase 2 ANN-I EFF model performance

The ANN-I model does not follow the data trend with a high accuracy, resulting in horizontal predictions on most of the data points. This is also an indication of undertraining, resulting in inaccurate model predictions. The validation data contains some extreme EFF values that could have caused the validation error to increase with an increase in training, resulting in the short training phase. The EFF main effects model shows an increase in the validation prediction accuracy, fitting most of the validation data points. The ANN-M model follows the trend with a higher accuracy despite its moderate error values. There are still some validation points that are predicted with a near constant value. The addition of independent variables increased the penalisation from the validation check, resulting in the lower performance from the ANN-I model. The ANN-M model is therefore preferred due to its simplicity and increased performance over the ANN-I model.

The product mean particle size (PPS) ANN model with all the variables have an increased CMD on the modelling, validation and overall datasets compared to the main effects model. The ANN-I model has a 9% difference between the CMD and ACD values indicating that some variables are unnecessary. The removal of these variables will, however, lower the performance of the model. The ANN-M model only trained for 3 epochs that might not be sufficient to ensure proper training. The training error decreased with further training as shown in Figure C-7, however, the validation error increased with further training beyond the third epoch as shown in Figure C-8, triggering the validation stop. Both figures start at the second epoch in order to observe the decrease and increase in error. From these figures, it is concluded that the validation data caused the training to stop early.

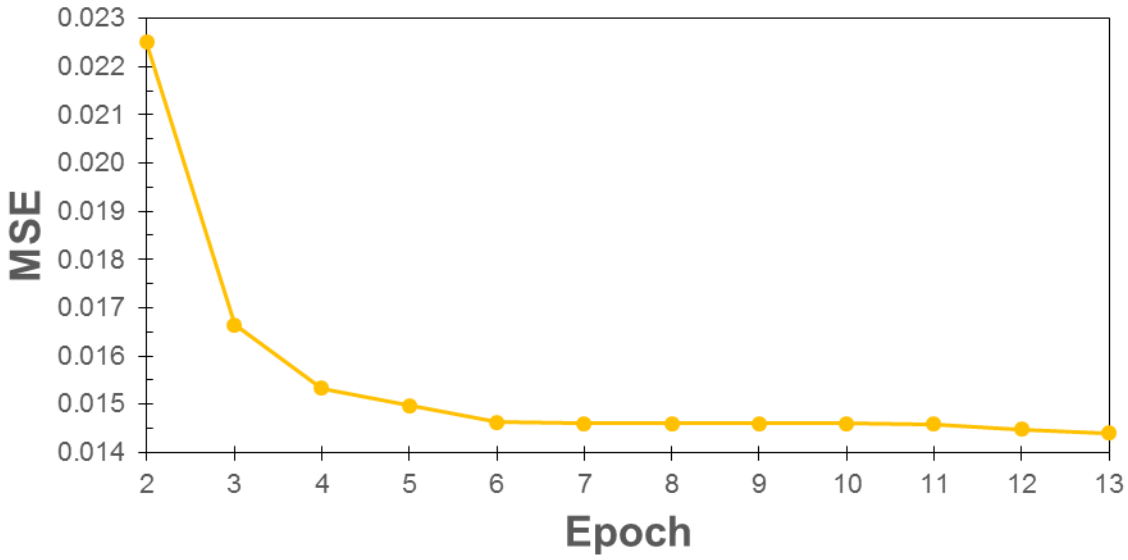


Figure C-7: Phase 2 ANN-M PPS model training error

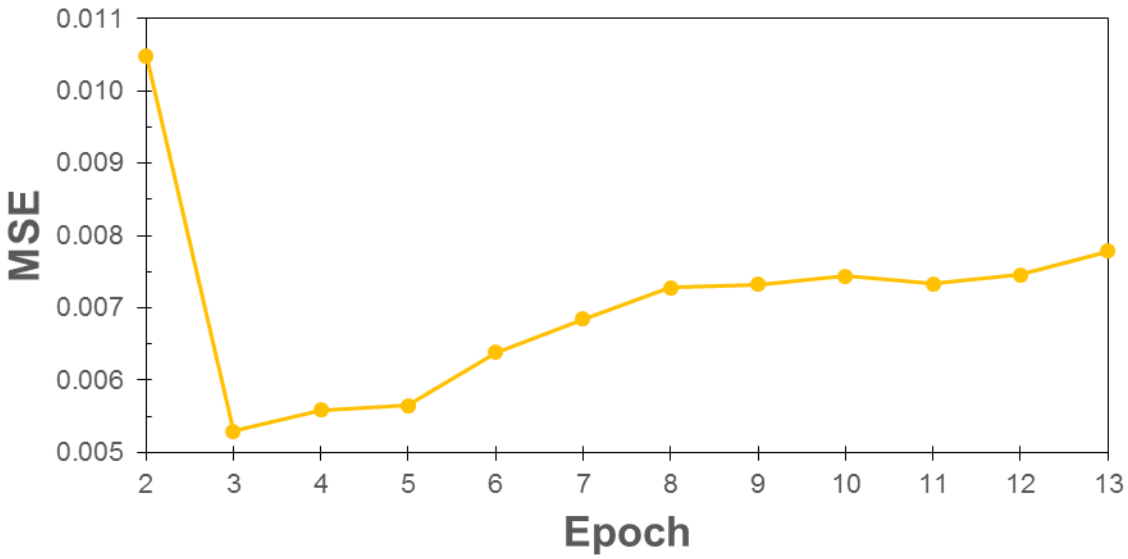


Figure C-8: Phase 2 ANN-M PPS model validation error

Both models have the same SEE and MAE values indicating that both models have the same amount of prediction deviation. The performances from both models are shown in Figure C-9 and Figure C-10. Both models follow the trend of the data with a low accuracy, struggling to fit the extreme points, e.g. data points 34 and 45. The undertraining due to the early training stopping of the ANN-M model is visible at points 1 to 18 and 58 to 64. The ANN-I model follows the trend better with a higher accuracy compared to the ANN-M model. Both models are concluded fairly inaccurate, underperforming on both the modelling and validation data.

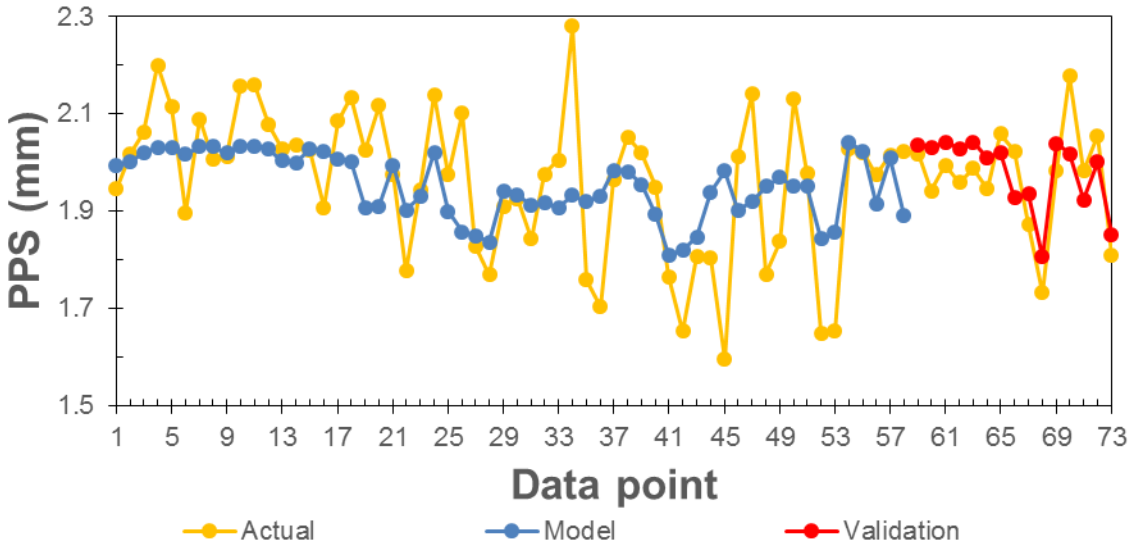


Figure C-9: Phase 2 ANN-M PPS model performance

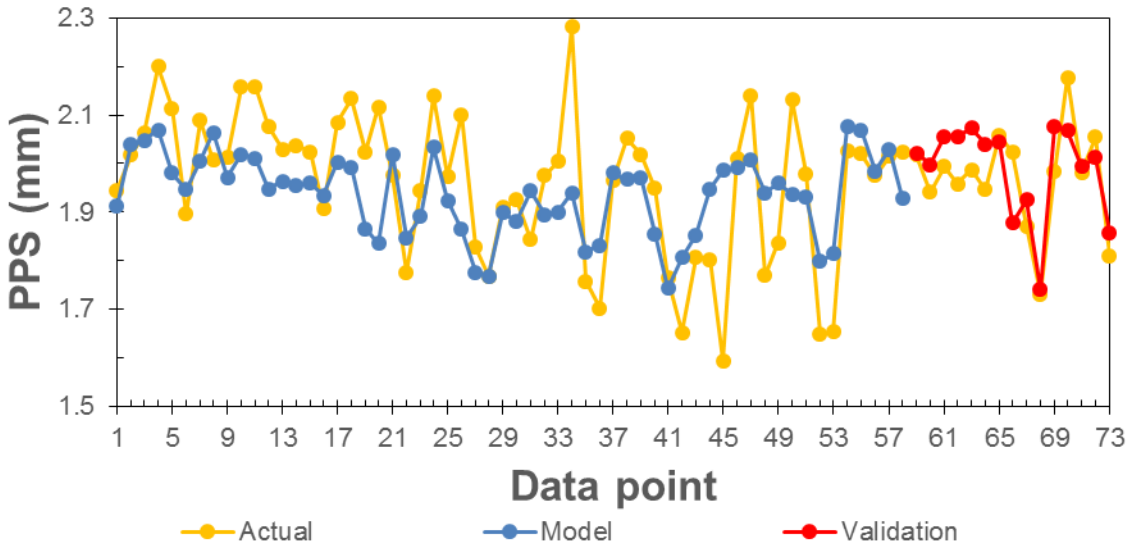


Figure C-10: Phase 2 ANN-I PPS model performance

There were no statistical significant main effects MLR models for the granulator mean particle size (GPS) and product porosity (POR) variables, resulting in only ANN-I models for both the GPS and POR variables. Both models have low CMD values for the modelling, validation and overall datasets. There is a significant difference between the CMD and the ACD value for the GPS and POR models. This indicates that some variables are unnecessary and might decrease the performance of the models. The POR model has a moderate SEE and MAE value, indicating deviations between the predicted and actual values. These models are considered inaccurate to be used for prediction purposes due to their low CMD and high error values.

The product circularity (PCR) ANN mode developed with all the variables has a 33% increase in the model CMD value compared to the main effects ANN model. The validation of the ANN-I model is moderate with a regression value of 51% higher than the ANN-M model. The model also has a moderate SEE and low MAE value indicating a deviation in the predicted and actual values. The model's performance increased significantly with the addition of independent variables, followed by an increase in the difference between the ACD and modelling CMD values. Both models show difficulty in modelling and predicting values far from the mean as shown in Figure C-11 and Figure C-12.

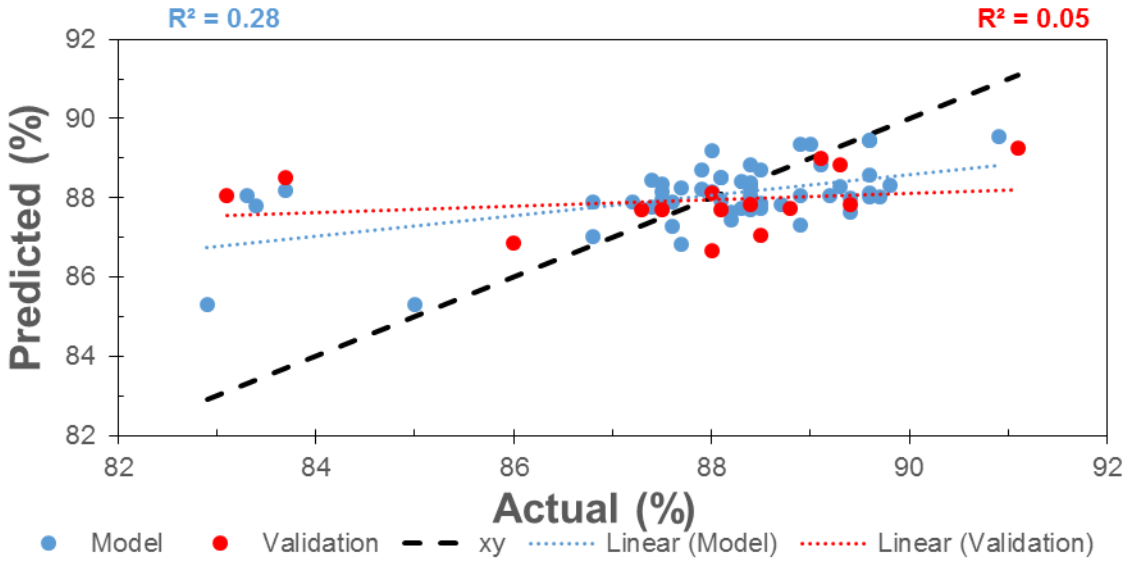


Figure C-11: Phase 2 ANN-M PCR model regression

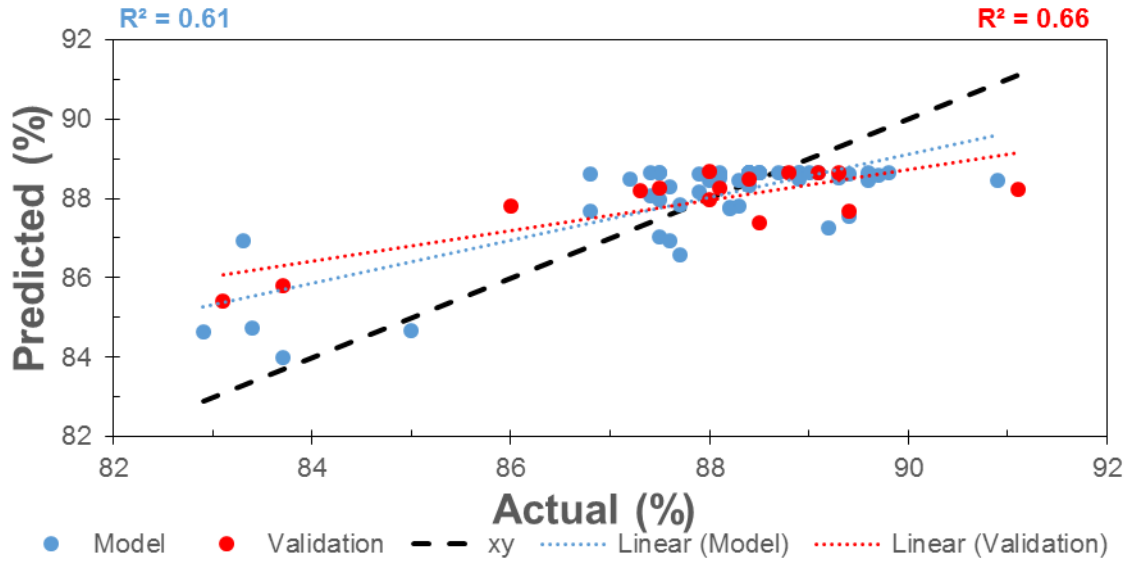


Figure C-12: Phase 2 ANN-I PCR model regression

The PCR ANN model with all the variables included performs better, but still cannot fit the dataset accurately. Some of the predicted values are oriented in a horizontal fashion indicating insufficient training. A few validation data points are situated at the lower end of the PCR value range and could be the cause to the insufficient training due to early training stopping. Both models are considered inaccurate to be used for prediction purposes.

The product particle slope (PSL) ANN models did not perform well on the modelling data but has a high CMD on the validation data. This is due to more validation data points distributed closer to the mean of the modelling data as shown in Figure C-13 and Figure C-14.

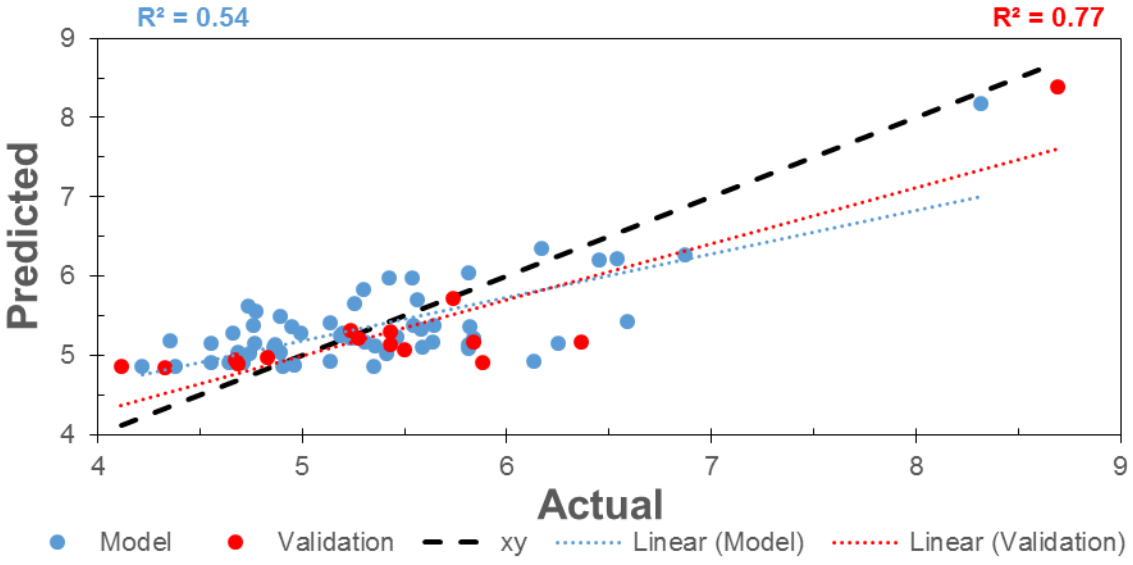


Figure C-13: Phase 2 ANN-M PSL model regression

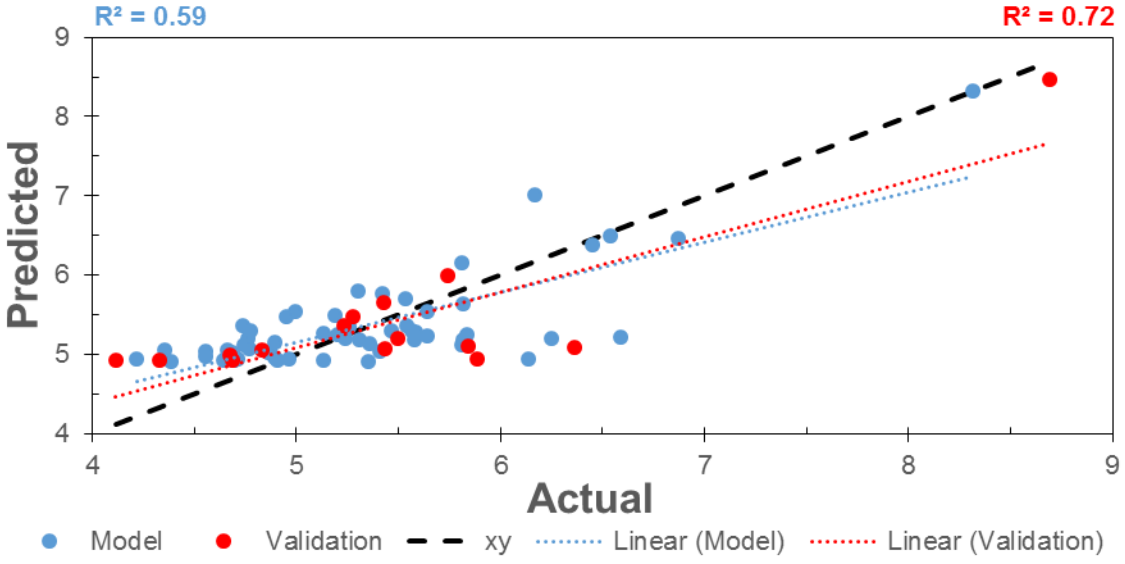


Figure C-14: Phase 2 ANN-I PSL model regression

There is one modelling and validation data point located at the extreme end of the PSL variable range. These points influence the CMD values greatly, skewing the regression line and resulting in the high regression values that do not fully represent the models accuracy and prediction capability. Both models have low SEE and MAE values indicating low deviations between the actual and predicted data points, indicating that only a few data points are distributed far from the regression line, as seen from the figures. Both models show almost the same fit with a relatively low accuracy, resulting in inaccurate models.

The granulator particle slope (GSL) main effects ANN model performed well with an 80% model CMD and 90% validation CMD along with low SEE and MAE values. Both models predict and follow the data trend well as shown in Figure C-15 and Figure C-16

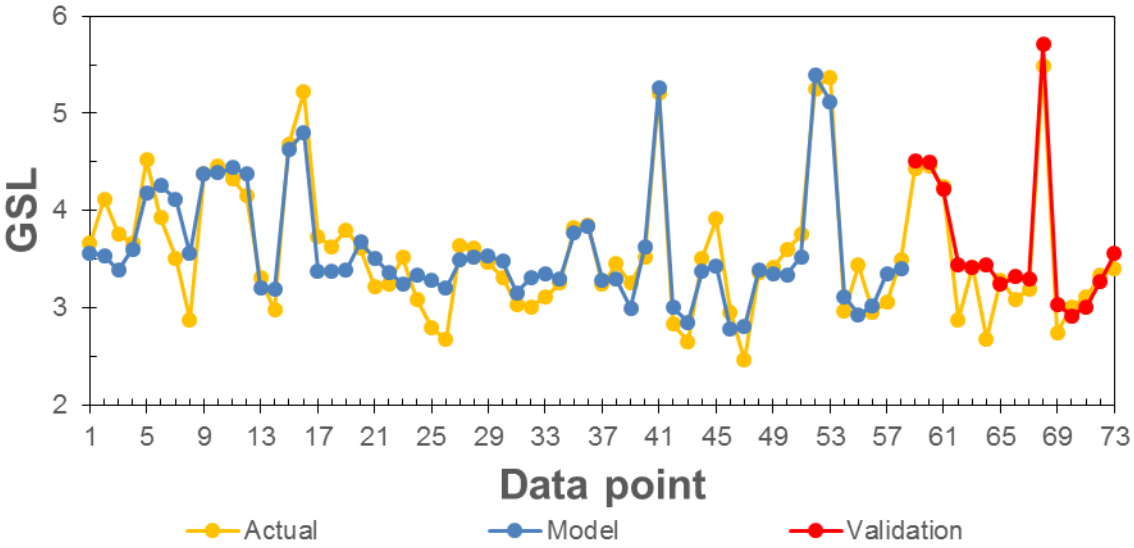


Figure C-15: Phase 2 ANN-M GSL model performance

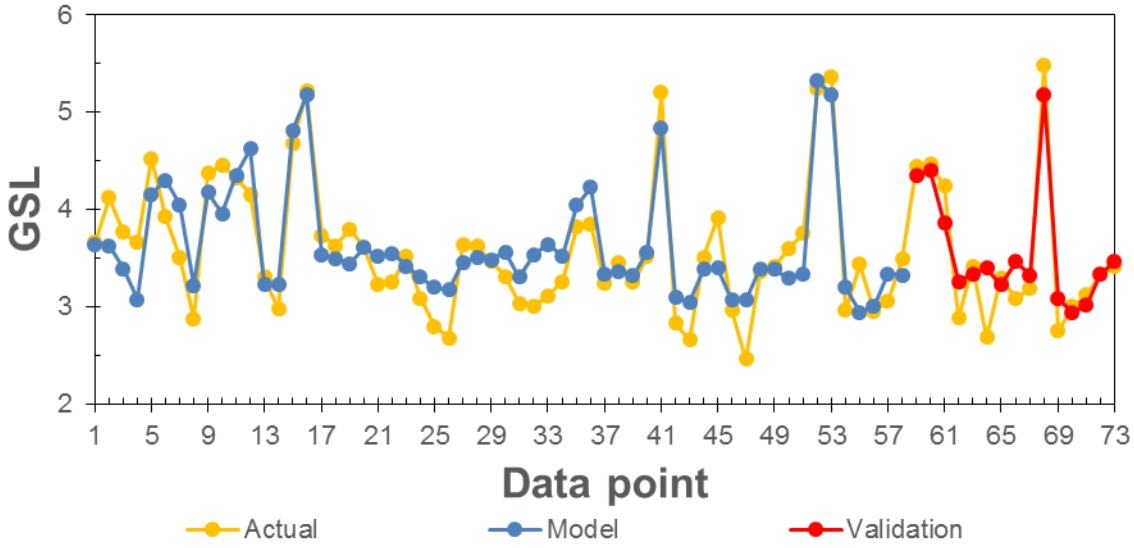


Figure C-16: Phase 2 ANN-I GSL model performance

The ANN-M model fits most of the data with a higher accuracy compared to the ANN-I model, confirming the lower SEE value obtained. The ANN-M model trained for 8 epochs longer, resulting in the increased accuracy. Both models struggle to predict the values at data points 2, 24 to 26 and 64. The ANN-M model follows the data trend better and fits the validation data with a higher accuracy, making it the preferred model for prediction.

It is concluded that the RUR, RER, EFF and GSL models followed the data trend with a higher accuracy, resulting in an adequate performance for prediction purposes. The ANN-M models performed well especially in the cases of the RUR, EFF and GSL models. Using the main effects MLR technique to determine the statistical significant variables helped in developing smaller, more accurate neural networks. This also adds some significance to the black-box approach by determining the inputs variables for the neural network statistically.

The number of weights of a neural network increases by adding more input variables. An ANN with more weights has a tendency to over fit or become stuck in a local minimum (Hill & Lewicki 2006, p.430). The added variables and weights also resulted in larger error values during the validation prediction phase. These larger error values resulted in an increase in the validation error resulting in a shorter training period due leading to undertrained ANNs. This was visible due to the near horizontal predicted values from the ANN models. This effect did not have a great affect the ANN-M models due to the smaller number of weights that needed to be trained.

The PPS, GPS, POR, PCR and PSL neural network models did not performed sufficiently. Larger neural network topographies or different initial weight values might increase the accuracies, however the purpose was to perform a qualitatively comparison between the modelling techniques.

Appendix D - Results

Spearman's rho results

The Spearman's rho matrix results for phase 1 are summarised in Table D-1. The table is divided into four quadrants with only the coloured three quadrants being used since the matrix is a reflection around the diagonal. The p-values for the rho values are summarised in Table D-2.

Table D-1: Spearman's rho matrix for phase 1

	FAF	FAT	SLF	SLT	SLC	SPS	SSL	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
FAF	1.00	-0.31	0.39	0.23	0.01	0.40	0.13	0.40	-0.45	0.40	0.13	0.32	0.09	0.08	-0.14	0.27
FAT	-0.31	1.00	-0.33	-0.02	-0.13	0.08	0.34	-0.39	0.48	-0.37	0.06	0.02	-0.25	0.49	0.22	0.14
SLF	0.39	-0.33	1.00	0.32	-0.28	0.42	-0.23	0.96	-0.76	0.51	0.29	0.66	0.17	-0.29	-0.77	-0.46
SLT	0.23	-0.02	0.32	1.00	0.15	0.08	-0.09	0.31	-0.31	0.69	0.06	0.06	-0.14	-0.13	-0.04	-0.18
SLC	0.01	-0.13	-0.28	0.15	1.00	-0.14	0.15	-0.26	-0.06	0.35	0.09	-0.23	-0.02	-0.37	0.21	0.01
SPS	0.40	0.08	0.42	0.08	-0.14	1.00	0.51	0.35	-0.52	0.08	0.75	0.72	-0.13	0.16	-0.30	0.02
SSL	0.13	0.34	-0.23	-0.09	0.15	0.51	1.00	-0.26	0.15	-0.28	0.33	0.20	-0.18	0.41	0.30	0.34
RUR	0.40	-0.39	0.96	0.31	-0.26	0.35	-0.26	1.00	-0.69	0.60	0.18	0.58	0.12	-0.30	-0.75	-0.46
RER	-0.45	0.48	-0.76	-0.31	-0.06	-0.52	0.15	-0.69	1.00	-0.49	-0.52	-0.52	-0.18	0.43	0.61	0.39
EFF	0.40	-0.37	0.51	0.69	0.35	0.08	-0.28	0.60	-0.49	1.00	0.07	0.14	0.05	-0.50	-0.30	-0.38
PPS	0.13	0.06	0.29	0.06	0.09	0.75	0.33	0.18	-0.52	0.07	1.00	0.70	-0.21	-0.05	-0.21	-0.10
GPS	0.32	0.02	0.66	0.06	-0.23	0.72	0.20	0.58	-0.52	0.14	0.70	1.00	-0.07	0.06	-0.42	-0.09
POR	0.09	-0.25	0.17	-0.14	-0.02	-0.13	-0.18	0.12	-0.18	0.05	-0.21	-0.07	1.00	-0.12	-0.17	-0.15
PCR	0.08	0.49	-0.29	-0.13	-0.37	0.16	0.41	-0.30	0.43	-0.50	-0.05	0.06	-0.12	1.00	0.32	0.47
PSL	-0.14	0.22	-0.77	-0.04	0.21	-0.30	0.30	-0.75	0.61	-0.30	-0.21	-0.42	-0.17	0.32	1.00	0.49
GSL	0.27	0.14	-0.46	-0.18	0.01	0.02	0.34	-0.46	0.39	-0.38	-0.10	-0.09	-0.15	0.47	0.49	1.00

Table D-2: Spearman's rho p-values for phase 1

	FAF	FAT	SLF	SLT	SLC	SPS	SSL	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
FAF		.07	.02	.19	.96	.02	.46	.02	.01	.02	.45	.06	.61	.64	.42	.12
FAT	.07		.05	.90	.46	.63	.05	.02	.00	.03	.75	.93	.14	.00	.20	.41
SLF	.02	.05		.06	.11	.01	.18	.00	.00	.00	.09	.00	.33	.09	.00	.01
SLT	.19	.90	.06		.38	.66	.60	.07	.07	.00	.73	.72	.41	.45	.83	.30
SLC	.96	.46	.11	.38		.43	.40	.14	.71	.04	.59	.18	.90	.03	.22	.96
SPS	.02	.63	.01	.66	.43		.00	.04	.00	.64	.00	.00	.44	.35	.08	.93
SSL	.46	.05	.18	.60	.40	.00		.13	.39	.10	.05	.24	.31	.01	.08	.04
RUR	.02	.02	.00	.07	.14	.04	.13		.00	.00	.30	.00	.49	.08	.00	.00
RER	.01	.00	.00	.07	.71	.00	.39	.00		.00	.00	.00	.31	.01	.00	.02
EFF	.02	.03	.00	.00	.04	.64	.10	.00	.00		.71	.43	.78	.00	.08	.02
PPS	.45	.75	.09	.73	.59	.00	.05	.30	.00	.71		.00	.22	.77	.24	.57
GPS	.06	.93	.00	.72	.18	.00	.24	.00	.00	.43	.00		.69	.73	.01	.59
POR	.61	.14	.33	.41	.90	.44	.31	.49	.31	.78	.22	.69		.50	.32	.39
PCR	.64	.00	.09	.45	.03	.35	.01	.08	.01	.00	.77	.73	.50		.06	.00
PSL	.42	.20	.00	.83	.22	.08	.08	.00	.00	.08	.24	.01	.32	.06		.00
GSL	.12	.41	.01	.30	.96	.93	.04	.00	.02	.02	.57	.59	.39	.00	.00	

The Spearman's rho matrix results for phase 2 are summarised in Table D-3 with the p-values for each rho value summarised in Table D-4.

Table D-3: Spearman's rho matrix for phase 2

	FAF	FAT	SLF	SLT	SLC	SPS	SSL	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
FAF	1.00	-0.07	-0.10	0.04	0.05	-0.02	-0.18	-0.06	0.00	-0.05	-0.16	-0.08	-0.03	-0.10	-0.28	0.01
FAT	-0.07	1.00	-0.16	-0.04	0.09	-0.26	0.51	-0.14	0.21	-0.09	-0.27	-0.09	0.22	0.01	-0.04	0.55
SLF	-0.10	-0.16	1.00	-0.05	-0.21	0.44	-0.22	0.74	-0.26	-0.04	0.10	0.16	0.14	0.01	-0.39	-0.39
SLT	0.04	-0.04	-0.05	1.00	0.75	-0.36	-0.02	0.17	-0.55	0.71	0.42	0.12	-0.23	-0.63	-0.15	0.35
SLC	0.05	0.09	-0.21	0.75	1.00	-0.44	-0.04	0.12	-0.41	0.72	0.28	-0.09	0.03	-0.49	-0.21	0.20
SPS	-0.02	-0.26	0.44	-0.36	-0.44	1.00	-0.09	0.30	0.13	-0.26	-0.12	0.26	-0.15	0.17	-0.14	-0.36
SSL	-0.18	0.51	-0.22	-0.02	-0.04	-0.09	1.00	-0.05	0.04	-0.08	-0.37	-0.13	0.17	-0.09	0.17	0.47
RUR	-0.06	-0.14	0.74	0.17	0.12	0.30	-0.05	1.00	-0.56	0.31	0.19	0.18	0.34	-0.20	-0.40	-0.34
RER	0.00	0.21	-0.26	-0.55	-0.41	0.13	0.04	-0.56	1.00	-0.63	-0.37	-0.11	-0.07	0.39	0.31	-0.01
EFF	-0.05	-0.09	-0.04	0.71	0.72	-0.26	-0.08	0.31	-0.63	1.00	0.36	0.18	0.01	-0.53	-0.21	0.20
PPS	-0.16	-0.27	0.10	0.42	0.28	-0.12	-0.37	0.19	-0.37	0.36	1.00	0.43	-0.21	-0.29	-0.03	-0.07
GPS	-0.08	-0.09	0.16	0.12	-0.09	0.26	-0.13	0.18	-0.11	0.18	0.43	1.00	-0.28	-0.11	0.15	0.08
POR	-0.03	0.22	0.14	-0.23	0.03	-0.15	0.17	0.34	-0.07	0.01	-0.21	-0.28	1.00	0.11	-0.17	-0.19
PCR	-0.10	0.01	0.01	-0.63	-0.49	0.17	-0.09	-0.20	0.39	-0.53	-0.29	-0.11	0.11	1.00	0.08	-0.33
PSL	-0.28	-0.04	-0.39	-0.15	-0.21	-0.14	0.17	-0.40	0.31	-0.21	-0.03	0.15	-0.17	0.08	1.00	0.22
GSL	0.01	0.55	-0.39	0.35	0.20	-0.36	0.47	-0.34	-0.01	0.20	-0.07	0.08	-0.19	-0.33	0.22	1.00

Table D-4: Spearman's rho p-values for phase 2

	FAF	FAT	SLF	SLT	SLC	SPS	SSL	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
FAF		.60	.46	.79	.70	.86	.17	.66	.99	.69	.22	.54	.80	.44	.03	.96
FAT	.60		.22	.75	.51	.05	.00	.28	.11	.50	.04	.49	.09	.93	.79	.00
SLF	.46	.22		.72	.12	.00	.10	.00	.05	.78	.45	.23	.29	.95	.00	.00
SLT	.79	.75	.72		.00	.01	.86	.21	.00	.00	.00	.38	.08	.00	.27	.01
SLC	.70	.51	.12	.00		.00	.76	.36	.00	.00	.03	.49	.82	.00	.11	.14
SPS	.86	.05	.00	.01	.00		.48	.02	.32	.05	.36	.05	.27	.21	.28	.01
SSL	.17	.00	.10	.86	.76	.48		.72	.79	.57	.00	.33	.20	.51	.21	.00
RUR	.66	.28	.00	.21	.36	.02	.72		.00	.02	.15	.19	.01	.13	.00	.01
RER	.99	.11	.05	.00	.00	.32	.79	.00		.00	.00	.42	.60	.00	.02	.94
EFF	.69	.50	.78	.00	.00	.05	.57	.02	.00		.01	.17	.94	.00	.11	.12
PPS	.22	.04	.45	.00	.03	.36	.00	.15	.00	.01		.00	.11	.03	.84	.58
GPS	.54	.49	.23	.38	.49	.05	.33	.19	.42	.17	.00		.04	.40	.27	.54
POR	.80	.09	.29	.08	.82	.27	.20	.01	.60	.94	.11	.04		.40	.19	.16
PCR	.44	.93	.95	.00	.00	.21	.51	.13	.00	.00	.03	.40	.40		.53	.01
PSL	.03	.79	.00	.27	.11	.28	.21	.00	.02	.11	.84	.27	.19	.53		.09
GSL	.96	.00	.00	.01	.14	.01	.00	.01	.94	.12	.58	.54	.16	.01	.09	

Multiple linear regression results

The statistical significant coefficients for the main effects multiple linear regression (MLR) model are summarised in Table D-5. The general equation for each MLR main effects model is given in Equation D-1.

$$\text{Quality variable} = a + b_1\text{FAF} + b_2\text{FAT} + b_3\text{SLF} + b_4\text{SLT} + b_5\text{SLC} + b_6\text{SPS} + b_7\text{SSL} \quad \textbf{Equation D-1}$$

Table D-5: MLR main effects model coefficients for phase 1

	RUR	RER	EFF	PPS	GPS	PCR	PSL	GSL
Constant	-10.76	38.91	-1.08	1.21	0.67	0.83	11.11	2.17
FAF	-	-2E-04	-	-3E-05	-	2E-06	-	3E-04
FAT	-0.01	0.01	-4E-04	-	-	2E-04	-0.02	-
SLF	0.62	-	-	-	0.02	-	-0.34	-0.20
SLT	0.03	-0.10	4E-03	-	-	-	-	-
SLC	0.09	-0.17	0.01	-	-	-	-	-
SPS	-	-4.22	-	0.75	0.45	-	-2.55	-
SSL	-	-	-	-	-	5E-03	0.76	-

Table D-6 summarises the statistical significant coefficients for the interaction effects MLR model for the product porosity (POR). The general equation for an interaction effects MLR model is given in Equation D-2.

$$\begin{aligned}\text{Quality variable} = {} & a + b_1\text{FAF} + b_2\text{FAT} + b_3\text{SLF} + b_4\text{SLT} + b_5\text{SLC} + b_6\text{SPS} + b_7\text{SSL} + b_8\text{FAF} * \text{FAT} + b_9\text{FAF} * \text{SLF} + \\ & b_{10}\text{FAF} * \text{SLT} + b_{11}\text{FAF} * \text{SLC} + b_{12}\text{FAF} * \text{SPS} + \\ & b_{13}\text{FAF} * \text{SSL} + b_{14}\text{FAT} * \text{SLF} + b_{15}\text{FAT} * \text{SLT} + \\ & b_{16}\text{FAT} * \text{SLC} + b_{17}\text{FAT} * \text{SPS} + b_{18}\text{FAT} * \text{SSL} + \\ & b_{19}\text{SLF} * \text{SLT} + b_{20}\text{SLF} * \text{SLC} + b_{21}\text{SLF} * \text{SPS} + \\ & b_{22}\text{SLF} * \text{SSL} + b_{23}\text{SLT} * \text{SLC} + b_{24}\text{SLT} * \text{SPS} + \\ & b_{25}\text{SLT} * \text{SSL} + b_{26}\text{SLC} * \text{SPS} + b_{27}\text{SLC} * \text{SSL} + \\ & b_{28}\text{SPS} * \text{SSL}\end{aligned} \quad \textbf{Equation D-2}$$

Table D-6: MLR interaction effects model coefficients for phase 1

	POR
Constant	280.46
FAT	-2.14
SLF	-27.73
FAF*FAT	-3E-04
FAF*SPS	0.02
FAT*SLF	0.25
FAT*SLT	0.03
SLT*SPS	-2.15

Table D-7summarises the statistical significant coefficients for the main effects MLR models for phase 2.

Table D-7: MLR main effects model coefficients for phase 2

	RUR	RER	EFF	PPS	PCR	PSL	GSL
Constant	-25.65	146.52	-481.29	1.41	130.21	25.99	-2.97
FAF	-	-	-	-5E-05	-	-	-
FAT	-	-	-	-	-	-	0.01
SLF	0.72	-0.77	1.69	-	-	-0.20	-0.12
SLT	-	-	-	9E-03	-	-	0.03
SLC	0.27	-1.45	5.89	-	-0.44	-0.21	-
SPS	-	-	-	-	-1.06	-0.22	0.27
SSL	-	-	-	-0.08	-	-	0.37

Table D-8 summarises the statistical significant coefficients for the interaction effects MLR models for phase 2.

Table D-8: MLR interaction effects model coefficients for phase 2

	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
Constant	1.51	325.18	-5541	2.19	-15.72	-364	753	22.39	13.92
FAF	-	-	-	-	-	-	-	-	-
FAT		-	-	-	-	-	-	-	-
SLF	-4.31	-29.67	-	-	-	-	-	-	-
SLT	-	-	38.81	-	-	8.42	-4.92	-	-
SLC	-	-3.44	60.20	-	0.18	-	-7.05	-0.17	-0.20
SPS	-	-	-	-	8.42	-196	-	-	-
SSL	-	-	-	-	0.65	-	-	-	-
FAF*FAT	-	-	-	-	-	-	-	-	-
FAF*SLF	-	-	-	-	-	-	-	-2E-05	-
FAF*SLT	-	-	-	-	-	-	-	-	-
FAF*SLC	-	-	-	-	-	-	-	-	-
FAF*SPS	-6E-05	-	-	-	-	-9E-04	-	-	2E-05
FAF*SSL	-	-	-	-7E06	-	1E-03	-	-	-
FAT*SLF	-	-	-	-	1E-03	-0.03	-	-	-
FAT*SLT	-	-	-	-	-	-0.01	3E-04	-	-
FAT*SLC	-	-	-	-	-	0.02	-	-	-
FAT*SPS	-	-	-	-	-	-	-0.01	-	-
FAT*SSL	-2E-03	0.02	-	-	-4E-03	-	-	-	0.01
SLF*SLT	-	-	-		-1E-03	-	-	-	-
SLF*SLC	0.05	0.33	-	-	-	-	-	-	-
SLF*SPS	0.14	-	-	-	-	-	0.12	-	-
SLF*SSL	0.09	-0.71	-	-	-	1.35	-	-	-0.06
SLT*SLC	-	-	-0.42	-	-	-0.06	0.05	-	6E-04
SLT*SPS	-	-	-	0.01	0.01	-0.86	-	-	-
SLT*SSL	-	-	-	-	-	-	-	-	-
SLC*SPS	-	-	-	-0.01	-0.10	3.56	-	-	-
SLC*SSL	-	-	-	-	-	-0.19	-	-	-
SPS*SSL	-	-	-	-	-0.10	-1.24	-	-	-

Appendix E - Data

The dataset used for the first phase is summarised in Table E-1.

Table E-1: Phase 1 dataset

Number	FAF	FAT	SLF	SLT	SLC	SPS	SSL	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
1	12200	113.12	9.30	135.32	92.04	1.46	3.83	6.28	3.15	76.60	1.92	1.59	21.86	88	5.16	4.09
2	12200	112.98	9.59	135.65	92.03	1.50	3.88	6.39	3.16	76.33	1.92	1.65	25.84	88	5.38	3.70
3	12400	113.04	9.68	135.85	92.03	1.50	3.78	6.59	3.08	77.38	1.86	1.52	27.17	89	5.14	3.52
4	12400	112.87	9.39	136.06	96.00	1.46	3.83	6.48	3.06	80.50	1.92	1.58	28.49	89	4.83	3.50
5	12280	89.62	8.74	135.93	91.99	1.63	2.87	6.06	2.99	78.57	1.92	1.45	25.70	88	4.68	4.08
6	12000	90.05	8.39	135.65	92.07	1.46	3.56	5.76	3.04	77.85	2.05	1.56	24.55	88	5.16	3.34
7	12300	90.27	8.79	135.67	91.96	1.47	3.70	6.18	2.99	78.81	2.07	1.54	26.33	88	6.51	3.38
8	12300	89.63	8.84	135.75	91.93	1.44	3.58	6.14	3.00	78.75	1.90	1.50	27.05	88	5.50	2.98
9	11200	113.50	6.79	130.10	90.65	1.25	3.02	4.09	5.36	68.40	1.87	1.50	25.39	89	6.80	5.15
10	11100	113.03	5.52	130.58	90.62	1.30	3.66	3.56	5.25	70.21	1.80	1.35	27.53	89	6.86	3.91
11	11271	112.84	5.66	130.24	90.64	1.36	3.84	3.31	5.46	69.33	1.77	1.43	26.22	89	6.72	4.36
12	11471	112.12	6.62	122.56	91.41	1.42	4.01	4.19	5.41	70.82	1.88	1.43	24.71	89	6.32	3.92
13	12200	112.63	6.66	122.55	91.40	1.38	3.55	4.23	5.33	71.34	1.88	1.53	25.15	89	5.44	3.94
14	11440	112.59	6.75	122.54	91.52	1.42	3.56	4.36	5.27	71.64	1.92	1.58	29.07	89	6.10	4.06
15	11910	113.08	6.74	122.53	91.42	1.50	3.79	4.28	5.23	71.73	1.91	1.51	25.10	89	5.57	3.57
16	10010	136.02	7.29	127.16	90.61	1.51	3.89	4.41	5.32	70.60	1.91	1.56	27.06	89	4.66	3.86
17	10331	135.40	7.24	128.46	90.74	1.51	4.01	4.64	5.30	71.05	2.11	1.57	24.61	89	6.28	3.61
18	10516	134.60	7.16	128.30	90.67	1.28	3.35	4.64	5.23	71.27	1.84	1.35	24.73	89	4.92	3.40
19	10762	131.17	8.30	127.93	90.64	1.41	3.46	4.50	5.22	70.94	1.83	1.45	27.48	88	5.09	2.86
20	11709	111.50	6.34	125.11	92.29	1.47	3.93	3.92	4.08	70.32	1.93	1.47	17.14	88	6.14	3.93
21	11610	112.46	6.66	125.00	92.17	1.43	4.03	4.06	4.28	70.69	1.95	1.50	28.74	88	5.91	3.87
22	11763	112.77	6.50	124.88	92.20	1.48	3.75	4.04	4.40	71.26	1.95	1.50	27.28	89	6.14	3.89
23	11400	112.26	6.35	124.68	92.31	1.43	3.84	4.19	4.36	73.19	1.88	1.48	27.68	88	6.64	4.06

24	9544	135.41	4.56	130.64	92.31	1.40	3.87	2.97	5.26	74.28	1.94	1.39	13.39	88	5.97	3.81
25	9018	134.79	3.84	130.66	92.23	1.32	3.80	2.76	5.45	73.64	1.88	1.39	22.78	89	7.33	3.82
26	9318	134.52	3.27	130.69	92.17	1.28	3.79	2.10	5.63	74.53	1.87	1.32	26.03	88	6.88	3.83
27	8909	135.00	3.30	130.71	92.25	1.30	3.58	2.15	5.75	76.08	1.85	1.34	23.20	88	7.45	3.59
28	9421	83.29	8.81	130.40	92.04	1.29	2.55	6.61	5.23	75.60	1.83	1.51	15.39	87	5.10	3.28
29	9399	96.41	8.50	130.05	92.09	1.20	2.05	5.57	5.16	75.73	1.84	1.47	33.63	87	5.11	2.72
30	10558	92.12	8.19	129.78	92.06	1.17	1.96	5.54	5.18	75.43	1.83	1.39	30.92	87	4.87	2.87
31	10218	83.31	7.94	129.79	92.09	1.23	2.58	5.42	5.15	75.72	1.75	1.30	30.17	87	5.35	2.96
32	12010	112.57	5.97	129.81	92.17	1.20	2.82	4.02	5.31	75.60	1.69	1.29	30.22	88	6.41	3.99
33	12369	112.79	6.00	129.81	92.12	1.17	2.69	4.13	5.42	75.11	1.63	1.33	14.24	88	6.68	4.20
34	12000	112.41	6.26	129.82	92.09	1.32	4.12	4.17	5.47	75.00	1.61	1.34	24.88	89	6.74	4.04
35	12282	112.97	6.15	129.83	92.13	1.37	4.30	4.12	5.49	75.02	1.67	1.36	30.11	89	6.74	4.27

The second phase dataset is summarised in Table E-2. The open spaces indicate the removal of outliers that were identified, evaluated and removed. The modelling data consist of the first 58 samples with the last 15 data samples, 59-73, forming the validation dataset.

Table E-2: Phase 2 dataset

Number	FAF	FAT	SLF	SLT	SLC	SPS	SSL	RUR	RER	EFF	PPS	GPS	POR	PCR	PSL	GSL
1	11560	113.0	3.5	140.0	93.1	1.41	1.85	2.37	9.60	76.94	1.95	1.72		88.00	5.53	3.66
2	11355	115.3	10.0	140.8	89.3	1.98	2.02		9.33	72.47	2.02	1.62	34.01	88.00	4.89	4.12
3	11446	110.7	8.2	143.1	91.8	2.12	1.46	5.35	8.93	73.54	2.06	1.55		86.80	6.14	3.76
4	10556	101.0	9.7	143.7	93.7	1.39	1.77		8.03	78.11	2.20	1.60	32.74	88.10	4.91	3.66
5	10760	124.6	4.1	146.2	93.6	1.45	2.45	2.88	8.11	78.93	2.11	1.51	32.89	87.60	4.73	4.52
6	10432	124.8	4.0	143.9	93.7	1.86	3.22	2.76	8.38	79.08	1.90	1.55	33.49	88.20	6.59	3.93
7	9332	123.7	3.9	144.1	93.8	1.94	2.92	2.75	8.43	79.47	2.09	1.55	32.02	87.60	5.13	3.51
8	11521	90.1	8.0	148.0	93.7	1.40	1.84	5.46	6.10	77.52	2.01	1.37	38.70	87.20	4.55	2.88
9	10639	135.2	5.6	144.6	93.6	1.57	3.00	3.81	7.13	75.96	2.01	1.49	36.36	87.50	4.87	4.38
10	10535	134.9	5.6	146.4	93.6	1.35	2.09	3.73	7.23	75.84	2.16	1.48	35.54	87.50	4.35	4.45
11	11923	123.9	5.4	151.7	93.7	1.57	2.43	3.65	7.12	77.39	2.16	1.60	35.68	88.40	4.77	4.33
12	12783	124.1	3.9	151.5	93.7	1.55	2.17	2.71	7.22	76.49	2.08	1.58	35.97	88.30	5.25	4.15
13	12333	90.8	3.5	141.4	93.6	1.05	0.97	2.39	7.73	77.04	2.03	1.38	36.32	87.50	6.87	3.30
14	12604	90.0	3.5	141.7	93.6	1.10	0.97	2.36	7.92	76.54	2.04	1.26	35.84	87.70	6.45	2.98
15	12029	123.7	4.6	151.5	93.7	1.51	3.12	3.15	8.16	77.51	2.03	1.42	37.97	87.40	4.76	4.68
16	11953	124.0	4.6	151.5	93.7	1.64	3.75	3.17	8.13	78.07	1.91	1.57	36.28	88.20	5.58	5.22
17	10942	100.6	5.9	141.0	91.7	1.68	2.29	3.95	7.21	76.55	2.09	1.66	39.91	88.30	5.64	3.73
18	11265	100.1	5.8	141.0	91.9	1.54	2.15	3.93	7.30	77.06	2.13	1.65	40.08	88.10	5.46	3.63
19	11747	100.9	5.1	134.3	91.8	1.57	2.63	3.68	7.44	76.64	2.03	1.61	38.35	88.10	5.54	3.80
20	11784	101.0	3.6	134.4	92.0	1.77	2.53	2.60	7.66	77.30	2.12	1.55	39.89	87.90	5.30	3.61
21	11327	113.0	8.4	141.4	92.1	2.08	2.59	5.58	8.66	75.08	1.98	1.42	41.42	88.70	4.71	3.22
22	12855	113.1	5.4	136.3	92.5	1.89	2.26	3.81	8.49	77.82	1.78	1.48	39.65	88.50	4.56	3.25
23	12440	112.8	5.6	136.4	92.5	1.99	1.89	3.83	8.52	77.62	1.94	1.51	38.73	88.50	4.86	3.52
24	9984	87.6	5.2	141.1	92.6	1.60	2.41	3.65	8.49	78.38	2.14	1.48	37.39	87.50	5.19	3.09
25	12431	124.0	8.5	133.4	92.1	1.87	1.77	4.82	9.67	63.66	1.97	1.55	38.72	88.90	4.69	2.79
26	12469	123.2	8.2	130.4	92.2	1.82	2.21	4.71	9.44	64.74	2.10	1.42	38.47	88.40	4.69	2.68
27	11527	130.8	4.1	126.9	92.3	1.52	2.42	2.72	9.38	71.34	1.83	1.33	41.54	88.40	5.56	3.64

28	12026	131.5	4.5	127.1	92.3	1.55	2.61	2.88	9.51	71.58	1.77	1.37	41.14	88.40	4.89	3.62
29	9687	131.4	5.6	131.0	92.2	1.91	2.77	3.46	9.68	71.11	1.91	1.50	41.58	88.40	5.36	3.47
30	10125	133.5	5.6	130.9	91.6	1.83	2.49	3.55	9.71	70.41	1.93	1.56	41.46	88.40	5.24	3.31
31	11469	100.9	7.8	134.0	90.8	1.78	2.72	4.72	6.84	65.33	1.84	1.44	39.46	88.50	4.68	3.03
32	11583	101.9	5.9	134.6	90.9	1.76	2.64	4.12	6.89	66.13	1.98	1.48	38.62	87.90	5.20	3.01
33	10372	124.4	7.4	130.6	90.0	2.00	2.95	4.40	8.32	65.54	2.01	1.59	38.79	89.10	5.31	3.11
34	9914	124.3	7.5	130.8	90.0	1.98	2.71	4.46	8.45	65.06	2.28	1.65	38.92	88.40	6.25	3.26
35	11276	132.5	4.7	133.7	91.0	1.18	2.57	2.24	16.08	53.19	1.76	1.31	39.32	88.90	5.42	3.82
36	10708	127.5	4.6	134.0	90.7	1.32	3.02	2.10	17.06	51.11	1.70	1.33	38.99	89.00	5.81	3.85
37	8812	113.7	5.2	130.8	90.9	2.25	2.17	3.19	9.26	68.57	1.97	1.52	37.93	89.60	5.00	3.24
38	8999	113.4	5.0	130.9	90.9	2.13	2.16	3.22	9.29	70.71	2.05	1.57	36.66	89.30	5.82	3.45
39	9239	100.6	5.7	127.9	88.4	2.14	1.75	2.19	16.72	42.38	2.02	1.43	34.69	89.60	6.54	3.25
40	10013	100.3	4.6	127.9	89.5	1.68	2.84	1.99	17.12	43.48	1.95	1.40	32.52	89.60	6.17	3.52
41	11885	113.1	2.6	123.9	89.5	1.62	4.03	1.09	18.51	46.63	1.76	1.43	35.00	90.90	8.32	5.21
42	13415	89.7	5.6	123.8	89.8	2.78	1.07	1.21	18.02	39.33	1.65	1.48	34.38	89.70	4.66	2.83
43	12057	90.5	5.3	124.1	89.6	2.38	0.95	2.30	15.03	47.12	1.81	1.49	36.49	89.60	4.78	2.66
44	10793	115.1	7.5	134.4	91.6	1.71	2.79	5.53	5.19	83.49	1.80	1.30	39.82	87.40	4.69	3.51
45	10366	113.9	6.9	136.7	92.1	1.70	2.59	5.03	5.52	81.92	1.59	1.44	37.74	88.40	5.41	3.92
46	11798	90.8	8.5	131.6	90.5	2.86	1.80	5.84	6.70	77.19	2.01	1.72	35.53	89.40	5.13	2.96
47	11646	90.7	8.4	132.6	90.8	3.07	1.78	5.77	6.71	76.58	2.14	1.78	34.35	88.90	4.64	2.46
48	10071	114.5	6.0	132.9	91.8	1.98	2.72	3.35	8.39	63.42	1.77	1.40	36.66	89.60	5.81	3.36
49	9995	113.7	6.0	133.9	91.7	1.77	2.54	3.44	8.45	64.41	1.84	1.44	36.50	89.80	5.81	3.41
50	11010	90.2	4.6	134.5	91.7	2.02	2.19	2.86	8.27	69.27	2.13	1.48	34.32	89.20	4.95	3.60
51	10789	89.9	4.5	134.7	91.9	1.95	2.55	2.92	8.16	71.66	1.98	1.61	32.41	83.30	5.64	3.75
52	12528	133.8	5.7	132.6	91.2	5.47	3.59	3.41	8.37	66.79	1.65	1.39	38.18	82.90	4.38	5.25
53	13059	133.6	5.8	134.6	91.3	5.46	3.09	3.34	9.16	63.89	1.65	1.52	33.28	85.00	5.35	5.37
54	11227	67.77	8.08	152.46	93.71	2.49	2.03	5.49	7.10	78.29	2.03	1.36	36.37	87.70	4.22	2.96
55	9810	78.86	8.20	140.23	94.06	2.11	2.03	6.58	4.81	90.57	2.02	1.49	41.15	86.80	4.96	3.44
56	11772	65.62	6.62	133.89	91.75	2.06	2.22	3.95	7.68	64.61	1.98	1.42	35.85	89.40	4.74	2.96

57	10560	78.85	5.67	141.05	92.02	2.11	2.62	3.61	8.57	70.62	2.02	1.62	35.02	83.40	5.59	3.05
58	12405	68.64	6.10	134.04	89.98	2.57	2.37	3.88	9.40	70.09	2.02	1.87	34.69	83.70	5.84	3.49
59	10677	123.97	5.42	150.30	93.80	1.53	3.05	3.67	8.55	76.21	2.02	1.53		88.10	5.84	4.44
60	11448	123.88	5.55	150.05	93.62	1.49	2.98	3.71	8.52	76.69	1.94	1.57	35.19	88.40	6.36	4.46
61	10435	122.8	5.6	150.2	93.8	1.54	1.70	3.84	8.13	78.23	1.99	1.52	35.62	87.50	5.43	4.24
62	11783	90.5	7.4	147.5	93.7	1.58	1.80	4.90	6.11	78.22	1.96	1.32	38.73	87.30	4.67	2.88
63	10547	100.95	7.25	149.44	93.57	2.27	1.13	4.69	8.13	77.11	1.99	1.54	36.09	88.50	5.88	3.42
64	11043	113.6	8.6	141.4	92.1	2.16	2.17	5.70	8.39	75.93	1.95	1.60	40.64	88.80	4.69	2.68
65	10102	89.1	6.3	141.1	92.6	1.62	2.28	4.46	8.34	78.72	2.06	1.61	40.85	88.00	5.50	3.29
66	9559	123.92	4.92	127.99	90.47	1.68	2.29	3.04	9.76	68.18	2.02	1.39	37.88	89.10	5.74	3.09
67	8882	123.90	6.17	127.45	90.67	1.73	2.64	3.79	9.91	67.23	1.87	1.40	38.10	89.30	5.43	3.19
68	12376	113.0	2.0	124.4	90.1	1.62	4.42	0.93	19.13	46.91	1.73	1.42	35.50	91.10	8.69	5.49
69	10997	68.40	9.19	151.09	93.86	2.59	1.97	6.19	7.01	78.01	1.98	1.46	36.94	88.00	4.33	2.75
70	10213	78.79	8.66	140.32	94.15	2.25	2.11	6.89	4.89	92.66	2.18	1.38	40.97	86.00	4.12	3.00
71	11629	69.11	6.93	134.22	91.78	2.19	2.27	3.99	7.41	65.47	1.98	1.50	36.33	89.40	4.83	3.12
72	11046	78.91	5.25	140.72	91.93	1.93	2.36	3.33	8.63	70.36	2.06	1.57	34.00	83.10	5.27	3.33
73	13201	68.85	5.92	133.64	90.05	2.25	2.76	3.75	9.64	68.74	1.81	1.65	34.51	83.70	5.24	3.41