

THE UNIVERSITY OF NEWCASTLE

School of Engineering

# **Assessing Coal Properties and Their Effects on Coking Performance: A Data Mining Approach**

**Lauren North**

BE (Chemical, Hons 1) (*University of New South Wales*)

A thesis submitted in fulfilment of the requirements for the  
Degree of Doctorate of Philosophy in Chemical Engineering

May 2020

This research was supported by an Australian Government Research Training  
Program (RTP) Scholarship, and ACARP

# Abstract

Within the cokemaking industry, the ability to accurately predict the quality of coke produced from a variety of global coal basins is limited. Since the most recent review of coke quality prediction models, completed by Díez et al. (2002) on a selection of key models within the literature, very little improvement in prediction ability has been shown. In particular, the existing coke quality prediction models are deficient in their ability to link and expand upon reported fundamental coal behaviour. Where emerging data mining techniques have been applied for model development, often a clearly defined and technically sound analytical process has not been described. In this context, data mining as one step in the knowledge discovery process, has presented a unique opportunity to provide further insight into the behaviour of coking coals, and in particular, allowed for integration and extension of fundamental coking behaviour.

## Research Outcomes

Based on the conclusions of an extensive literature review considering both coking behaviour, and analysis of the methods of prediction of coke quality, a sub-model approach to prediction was developed. This sub-model approach addresses the some of the fundamental processes occurring within the formation of coke from the parent coals, whilst integrating data mining techniques. The following knowledge gaps were explored, with clearly defined frameworks for prediction developed:

### *Vitrinite reflectance distribution classification using self organising maps*

Vitrinite reflectance is perceived as an important parameter in coke quality, however in coal blending, utilising the average value is unsuitable as it does not reflect how two coals of vastly different reflectance have been blended.

Classification of the resulting distributions allows further insight into these blending decisions.

#### *Prediction of coal fusibility using a Sugeno fuzzy inference system*

Traditional prediction models consider fusibility of coal sub-components as one of the main factors in determining coke quality. However, the assumed proportion of fusing and non-fusing components is unreliable for Australian and other coals. Hence, accurate prediction of these fusing components for each respective coal ought to improve prediction of coke quality. Data quality issues associated with the comparatively small and biased data set were also explored. This section discusses how some of these issues can be addressed by using a modified data oversampling technique based on the Synthetic Minority Oversampling Technique (SMOTE).

#### *Approximation of coal mineral composition from coal ash chemistry using a genetic algorithm*

Coal ash chemistry is commonly used in the prediction of coke reactivity. However, the use of ash chemistry may be misleading as it does not directly reflect the mineral forms present in the coal which may have different influences on reactivity.

#### *Inference of coal mineral behaviour on coke reactivity using a support vector machine*

Whilst basicity index, derived from coal ash chemistry, is often applied in coke quality predictions, questions are raised over the relative importance of different minerals on coke reactivity, and whether the index is misleading. Using the approximated coal mineral composition, an improved understanding of coal reactivity behaviour is developed.

*Final prediction and assessment of individual sub-models on coke quality using a support vector regression*

Results show statistically significant improvement in predictive accuracy with the use of the vitrinite reflectance distribution terms, as well as the use of coal mineral composition.

The development and integration of these models in a clearly defined predictive framework presents a significant improvement to many traditional predictive models. Further, consistency with experimental studies was observed, with several areas for future work identified. This work has significant implications on not only the methods of prediction of coke quality, but also on the confirmation and integration of experimental findings.

# Certificate of Originality

I hereby certify that the work embodied in this thesis is my own work, conducted under normal supervision.

The thesis contains published scholarly work of which I am a co-author. For each such work a written statement, endorsed by the other authors, attesting to my contribution to the joint work has been included.

The thesis contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to the final version of my thesis being made available worldwide when deposited in the University's Digital Repository, subject to the provisions of the Copyright Act 1968, and any approved embargo.

Lauren North – PhD Candidate

Signature \_\_\_\_\_ Date 19<sup>th</sup> May 2020

# Acknowledgement of Authorship

I hereby certify that the work embodied in this thesis contains published paper/s/scholarly work of which I am a joint author. I have included as part of the thesis a written declaration endorsed in writing by my supervisor, attesting to my contribution to the joint publication/s/scholarly work.

By signing below I confirm that Lauren North contributed the primary intellectual input to the creation of and all analysis contained in the following papers and publications:

North LA, Blackmore KL, Nesbitt KV, Hockings K, Mahoney MR (2019) Understanding the impact of coal blending decisions on the prediction of coke quality: a data mining approach. *International Journal of Coal Science & Technology* 2:207-217 doi <https://doi.org/10.1007/s40789-018-0217-2>

North L, Blackmore K, Nesbitt K, Hockings K, Mahoney M (2018) Exploration of Coal Fusibility in the Cokemaking Process and the Links to Coke Quality. Paper presented at the 8th International Congress on Science and Technology of Ironmaking - ICSTI 2018, Vienna, Austria

North L, Blackmore K, Nesbitt K, Mahoney M (2018) Models of coke quality prediction and the relationships to input variables: A review *Fuel* 219:446-466 doi:<https://doi.org/10.1016/j.fuel.2018.01.062>

North L, Blackmore K, Nesbitt K, Mahoney M (2018) Methods of coke quality prediction: A review *Fuel* 219:426-445 doi:<https://doi.org/10.1016/j.fuel.2018.01.090>

North L, Blackmore K, Nesbitt K, Hockings K, Mahoney M (2017) A Novel Approach to Coke Strength Prediction Using Self Organizing Maps. Paper presented at the DMIN'17 - The 13th International Conference on Data Mining, Las Vegas, USA

Merrick Mahoney – Primary Supervisor

# Acknowledgements

First and foremost, I would like to thank my supervisors Merrick Mahoney, Karen Blackmore, and Keith Nesbitt. Words cannot articulate the degree of appreciation I have for this team. Without their patience, knowledge, discussions and encouragement, this thesis would not have been possible.

My gratitude goes to BHP and in particular Kim Hockings and the Coal Technical Marketing team. The insightful discussions and mentoring throughout this process have been invaluable.

Thanks also to ACARP and the cohort of the Technical Market Support researchers. The financial support and knowledge sharing enabled by this industry research program are gratefully acknowledged.

To my former colleagues at BlueScope, in particular, David Pinson, Sheng Chew, and Paul Zulli, I am grateful for the solid foundation in industrial research, and research writing. I've finally grown to appreciate the scrawls of red ink across my drafts.

To my colleagues at the University of Newcastle, in particular Hannah Lomas, Richard Roest, and Kim Colyvas. Thank you for your insightful feedback and discussions on your areas of specialty.

Finally, to my family and friends, thank you for your patience and support over these years.

# List of Publications

Parts of this thesis have been published in the following works:

North LA, Blackmore KL, Nesbitt KV, Hockings K, Mahoney MR (2019) Understanding the impact of coal blending decisions on the prediction of coke quality: a data mining approach. International Journal of Coal Science & Technology 2:207-217 doi <https://doi.org/10.1007/s40789-018-0217-2>

North L, Blackmore K, Nesbitt K, Hockings K, Mahoney M (2018) Exploration of Coal Fusibility in the Cokemaking Process and the Links to Coke Quality. Paper presented at the 8th International Congress on Science and Technology of Ironmaking - ICSTI 2018, Vienna, Austria

North L, Blackmore K, Nesbitt K, Mahoney M (2018) Models of coke quality prediction and the relationships to input variables: A review Fuel 219:446-466 doi:<https://doi.org/10.1016/j.fuel.2018.01.062>

North L, Blackmore K, Nesbitt K, Mahoney M (2018) Methods of coke quality prediction: A review Fuel 219:426-445 doi:<https://doi.org/10.1016/j.fuel.2018.01.090>

North L, Blackmore K, Nesbitt K, Hockings K, Mahoney M (2017) A Novel Approach to Coke Strength Prediction Using Self Organizing Maps. Paper presented at the DMIN'17 - The 13th International Conference on Data Mining, Las Vegas, USA



# Table of Contents

|                  |   |           |
|------------------|---|-----------|
| <b>Chapter 1</b> | <b>Introduction.....</b>  | <b>i</b>  |
| 1.1              | Background.....   | 1         |
| 1.2              | Thesis Structure .....  | 3         |
| 1.3              | Importance of Coal .....  | 4         |
| 1.3.1            | Coking Coal Formation .....                                       | 4         |
| 1.3.2            | The Coke Making Process .....                                     | 5         |
| 1.3.3            | Coal Quality Metrics .....  | 8         |
| 1.3.4            | Coal in the Blast Furnace.....                                    | 13        |
| 1.3.5            | Coke Quality Measures .....                                       | 14        |
| 1.3.6            | Prediction of Coke Quality .....                                  | 16        |
| 1.4              | Conclusion.....   | 16        |
| <b>Chapter 2</b> | <b>Relationship between coke quality and input variables.....</b> | <b>17</b> |
| 2.1              | Method of Review .....  | 17        |
| 2.2              | Early Models of Coke Quality Prediction .....                     | 18        |
| 2.2.1            | Ammosov, Schapiro and Gray.....                                   | 18        |
| 2.2.2            | Simonis G-Factor.....   | 23        |
| 2.2.3            | MOF Diagram.....  | 26        |
| 2.2.4            | Mechanistic Models.....   | 26        |
| 2.3              | CSR and CRI Prediction.....                                       | 28        |
| 2.3.1            | Prediction from Additive Behaviour .....                          | 28        |
| 2.3.2            | Relationship Between CSR and CRI .....                            | 30        |
| 2.3.3            | The Role of Rank.....   | 32        |
| 2.3.4            | The Role of Ash and Minerals.....                                 | 41        |
| 2.3.5            | The Role of Thermoplastic and Swelling Properties.....            | 48        |

|                  |   |           |
|------------------|---|-----------|
| 2.3.6            | The Role of Inerts .....  | 56        |
| 2.3.7            | The Role of Textures, Porosity, And Pore Structure .....                        | 57        |
| 2.3.8            | The Role of Operating Conditions.....   | 69        |
| 2.4              | Prediction of Other Coke Quality Parameters .....                               | 71        |
| 2.4.1            | ASTM Hardness and Stability .....   | 71        |
| 2.4.2            | Micum Indices .....   | 71        |
| 2.4.3            | Irsid and Japanese Drum Indices .....   | 74        |
| 2.5              | Summary and Conclusion.....   | 75        |
| <b>Chapter 3</b> | <b>Methods of coke quality prediction.....</b>                                  | <b>77</b> |
| 3.1              | An Introduction to Data Mining .....  | 77        |
| 3.2              | A Brief Introduction to the Knowledge Discovery Process .....                   | 78        |
| 3.3              | Method of Study .....   | 80        |
| 3.4              | Business Understanding .....  | 80        |
| 3.5              | Data Understanding and Sources.....   | 81        |
| 3.5.1            | Comparison Between the Use of Laboratory, Experimental and Industrial Data..... | 84        |
| 3.5.2            | Comparison Between Predictions of Blends and Single Coals .....                 | 85        |
| 3.5.3            | Controlled and Restricted Parameters .....                                      | 86        |
| 3.5.4            | Implications for Data Mining Techniques.....                                    | 94        |
| 3.6              | Data Preparation .....  | 95        |
| 3.6.1            | Data Cleaning .....   | 95        |
| 3.6.2            | Data Transformation.....  | 96        |
| 3.6.3            | Attribute Selection .....   | 99        |
| 3.6.4            | Final Data Properties .....   | 101       |
| 3.7              | Modelling Approaches .....  | 103       |
| 3.7.1            | Regression Approaches .....   | 103       |

|                  |  |            |
|------------------|--|------------|
| 3.7.2            | Artificial Neural Network Approaches.....  | 105        |
| 3.7.3            | Other Data Mining Techniques .....   | 115        |
| 3.8              | Evaluation.....  | 117        |
| 3.9              | Conclusion.....  | 117        |
| <b>Chapter 4</b> | <b>Conceptual Framework.....</b>   | <b>118</b> |
| 4.1              | Assessment of Knowledge Gaps .....   | 118        |
| 4.2              | Research Proposal .....  | 119        |
| 4.2.1            | Data Sources .....   | 120        |
| 4.2.2            | Method of Study .....  | 121        |
| 4.3              | Plan of Research .....   | 122        |
| 4.3.1            | Identified Sub-Models .....  | 122        |
| 4.4              | Structure of Following Chapters .....  | 123        |
| <b>Chapter 5</b> | <b>Part A - Vitrinite reflectance distributions and coal blending decision<br/>impacts on coke quality .....</b> | <b>124</b> |
| 5.1              | Chapter Summary .....  | 124        |
| 5.2              | Introduction .....   | 125        |
| 5.3              | A Summary of Relevant Data Mining Techniques .....   | 127        |
| 5.3.1            | Clustering for Pattern Recognition .....   | 128        |
| 5.4              | General Modelling Framework .....  | 130        |
| 5.4.1            | Data Cleaning .....  | 130        |
| 5.4.2            | Data Offsetting .....  | 130        |
| 5.4.3            | Data Conversion .....  | 131        |
| 5.4.4            | Determination of Modelling Parameters .....  | 131        |
| 5.4.5            | Generation of the Classes .....  | 131        |
| 5.4.6            | Determination of Model Robustness .....  | 132        |
| 5.4.7            | Fitting a Regression Equation.....   | 132        |

|   |   |            |
|---|---|------------|
| 5.5   | Modelling Results.....  | 132        |
| 5.5.1   | Introduction of Low CSR Cases in VDC Generation.....              | 133        |
| 5.5.2   | Alternate Rank Term, Volatile Matter.....                         | 134        |
| 5.6   | Interpretation of Vitrinite Reflectance Distribution Models ..... | 137        |
| <b>Chapter 5 Part B – Rule based extraction of vitrinite reflectance distributions</b><br>..... |   | <b>138</b> |
| 5.7   | Decision Tree Based Rule Extraction.....                          | 138        |
| 5.8   | An introduction to decision trees, and the C4.5 algorithm.....    | 138        |
| 5.9   | Implementation of decision trees.....                             | 139        |
| 5.9.1   | Data Manipulation and Attribute Creation .....                    | 140        |
| 5.9.2   | Data Exploration and Attribute Selection.....                     | 141        |
| 5.10  | Decision Tree Development .....                                   | 143        |
| 5.10.1  | Baseline .....  | 143        |
| 5.10.2  | CART Decision Tree .....  | 143        |
| 5.10.3  | C4.5 Decision Tree .....  | 143        |
| 5.11  | Tree Performance .....  | 143        |
| 5.11.1  | Performance Metrics.....  | 143        |
| 5.11.2  | ZeroR Baseline .....  | 145        |
| 5.11.3  | CART Decision Tree .....  | 145        |
| 5.11.4  | C4.5 / J48 Decision Tree .....                                    | 146        |
| 5.12  | Discussion on Optimum Algorithm.....                              | 148        |
| 5.13  | Optimisation of the C4.5 Parameters.....                          | 148        |
| 5.14  | Evaluation of Misclassified Instances .....                       | 150        |
| 5.15  | Validation of Model Performance .....                             | 151        |
| 5.16  | Conclusion.....   | 151        |
| <b>Chapter 6 Prediction of fused material in coke.....</b>                                      |   | <b>152</b> |

|       |  |     |
|-------|--|-----|
| 6.1   | Chapter Summary .....                                    | 152 |
| 6.2   | Introduction .....                                       | 153 |
| 6.3   | Literature review .....                                  | 154 |
| 6.3.1 | Historical approaches to fusibility representation ..... | 154 |
| 6.3.2 | Contributing factors to fusibility .....                 | 157 |
| 6.3.3 | A definition of fusibility .....                         | 158 |
| 6.4   | Synthesis, Hypothesis and Considerations .....           | 158 |
| 6.5   | Data Understanding .....                                 | 158 |
| 6.5.1 | Data Source.....   | 159 |
| 6.5.2 | Data Restrictions.....                                   | 160 |
| 6.5.3 | Measurement variation between categories of coals .....  | 160 |
| 6.5.4 | Measurement uncertainty .....                            | 162 |
| 6.5.5 | Imbalanced representation of coals .....                 | 162 |
| 6.6   | Data preparation .....                                   | 163 |
| 6.6.1 | Data cleaning .....                                      | 163 |
| 6.6.2 | Data transformation .....                                | 164 |
| 6.6.3 | Addressing the class imbalance .....                     | 164 |
| 6.6.4 | Managing uncertain variables and expert knowledge .....  | 165 |
| 6.6.5 | Attribute selection.....                                 | 166 |
| 6.7   | Model Framework .....                                    | 168 |
| 6.7.1 | SMOTE.....   | 170 |
| 6.7.2 | Fuzzy logic modelling .....                              | 170 |
| 6.7.3 | Tuning of modelling parameters.....                      | 170 |
| 6.8   | Model performance .....                                  | 174 |
| 6.8.1 | Fusibility prediction performance .....                  | 174 |
| 6.8.2 | Validation of behaviour of input parameters .....        | 175 |

|                  |   |            |
|------------------|---|------------|
| 6.9              | Examination of model features selected by data mining.....      | 177        |
| 6.9.1            | Impact of SMOTE .....   | 177        |
| 6.9.2            | Model limitations.....  | 178        |
| 6.10             | Preliminary evaluation of effect on coke quality .....          | 179        |
| 6.10.1           | Method of application.....                                      | 179        |
| 6.10.2           | Resulting predictions of coke quality .....                     | 180        |
| 6.11             | Conclusion.....   | 181        |
| <b>Chapter 7</b> | <b>Part A – Understanding the Modified Basicity Index .....</b> | <b>182</b> |
| 7.1              | Introduction .....  | 182        |
| 7.2              | Literature Review .....   | 183        |
| 7.2.1            | Measurement of Minerals and Ash Chemistry .....                 | 183        |
| 7.2.2            | Alternate methods of Mineral Determination.....                 | 185        |
| 7.2.3            | Minerals Identified in Coal.....                                | 185        |
| 7.2.4            | Effect of Minerals on Coke Reactivity .....                     | 186        |
| 7.3              | Synthesis, Hypothesis and Considerations .....                  | 188        |
| 7.4              | Data Understanding .....  | 188        |
| 7.4.1            | Data Source.....  | 188        |
| 7.4.2            | Data Restrictions.....  | 189        |
| 7.5              | Data Preparation .....  | 189        |
| 7.5.1            | Data Cleaning .....   | 189        |
| 7.5.2            | Data Transformations .....                                      | 190        |
| 7.5.3            | Imposed Data Assumptions .....                                  | 190        |
| 7.5.4            | Attribute Selection .....                                       | 191        |
| 7.6              | Model Framework .....   | 191        |
| 7.6.1            | System of Equations .....                                       | 193        |
| 7.6.2            | Genetic Algorithm Approaches .....                              | 194        |

|   |   |            |
|---|---|------------|
| 7.6.3   | Tuning of Model Parameters .....                  | 194        |
| 7.7   | Model Performance .....                           | 195        |
| 7.8   | Mineral Prediction Model Evaluation .....         | 198        |
| 7.8.1   | Application in Part B .....                       | 198        |
| 7.8.2   | What does the basicity index imply? .....         | 198        |
| <b>Chapter 7 Part B – Exploration of Mineral Behaviour Effects on Coke Quality</b><br>..... |   | <b>200</b> |
| 7.9   | Synthesis, Hypothesis and Considerations .....    | 200        |
| 7.10  | Data Understanding .....                          | 201        |
| 7.10.1  | Data Source.....                                  | 201        |
| 7.10.2  | Data Restrictions.....                            | 201        |
| 7.10.3  | Data Limitations .....                            | 201        |
| 7.11  | Data Preparation .....                            | 202        |
| 7.11.1  | Data Cleaning .....                               | 202        |
| 7.11.2  | Data Transformations .....                        | 204        |
| 7.11.3  | Attribute Selection .....                         | 204        |
| 7.12  | Model Framework .....                             | 211        |
| 7.12.1  | Support Vector Machines .....                     | 211        |
| 7.12.2  | Selection of Kernel Function.....                 | 213        |
| 7.12.3  | Tuning of Model Parameters .....                  | 213        |
| 7.13  | Model Performance .....                           | 215        |
| 7.13.1  | Predictive Performance.....                       | 215        |
| 7.13.2  | Validation of behaviour of input parameters ..... | 217        |
| 7.14  | Examination of Attribute Selection .....          | 217        |
| 7.15  | Mineral Prediction Model Evaluation .....         | 220        |
| 7.16  | Conclusion.....                                   | 222        |

|                  |   |            |
|------------------|---|------------|
| <b>Chapter 8</b> | <b>Integration and interpretation of sub-models within a coke quality prediction framework.....</b> | <b>223</b> |
| 8.1              | Introduction .....  | 223        |
| 8.2              | Literature Review .....   | 224        |
| 8.3              | Synthesis, Hypothesis and Considerations .....  | 224        |
| 8.4              | Data Understanding .....  | 226        |
| 8.4.1            | Data Source.....  | 226        |
| 8.4.2            | Data Limitations .....  | 226        |
| 8.4.3            | Data Restrictions.....  | 227        |
| 8.5              | Data Preparation .....  | 227        |
| 8.5.1            | Data Cleaning .....   | 227        |
| 8.5.2            | Data Transformations .....  | 229        |
| 8.5.3            | Imposed Data Assumptions .....  | 230        |
| 8.5.4            | Attribute Selection .....   | 230        |
| 8.6              | Model Framework .....   | 232        |
| 8.6.1            | Proposed models for consideration.....  | 233        |
| 8.6.2            | Tuning of Model Parameters .....  | 233        |
| 8.7              | Model Performance .....   | 233        |
| 8.7.1            | General Observations – Linear vs SVM.....   | 233        |
| 8.7.2            | General Observations – Fundamental Behaviour .....  | 234        |
| 8.8              | Evaluation of Models .....  | 237        |
| 8.8.1            | Evaluation of SOM Behaviour .....   | 237        |
| 8.8.2            | Evaluation of Fusibility Behaviour.....   | 242        |
| 8.8.3            | Evaluation of Mineral Terms.....  | 244        |
| 8.8.4            | Evaluation of Combination Models.....   | 247        |
| 8.8.5            | Evaluation of Model Series 7 – Terms selected by attribute selection.....                           | 248        |



|                   |   |            |
|-------------------|---|------------|
| 8.8.6             | Proposal of Best Model .....  | 249        |
| 8.8.7             | Limitations of Model .....  | 256        |
| 8.9               | Conclusion.....   | 257        |
| <b>Chapter 9</b>  | <b>Conclusion and Future Work.....</b>  | <b>259</b> |
| 9.1               | Introduction .....  | 259        |
| 9.2               | Existing predictions of coke quality .....  | 259        |
| 9.3               | Importance of relationship to fundamental coking behaviour.....                               | 260        |
| 9.4               | Accounting for blending decisions using vitrinite reflectance distributions ...               | 260        |
| 9.5               | Understanding of coal fusibility .....  | 262        |
| 9.6               | Understanding of the basicity index relationships with reference to mineral composition ..... | 263        |
| 9.7               | An improved model for coke quality prediction .....   | 264        |
| 9.8               | Areas for future study.....   | 265        |
| 9.9               | Summary.....  | 267        |
| <b>Chapter 10</b> | <b>References.....</b>  | <b>268</b> |
| <b>Appendix A</b> | <b>List of Symbols.....</b>   | <b>302</b> |
| <b>Appendix B</b> | <b>CSR and CRI correlations .....</b>   | <b>307</b> |
| <b>Appendix C</b> | <b>Correlations for other metrics of coke quality .....</b>                                   | <b>327</b> |
| <b>Appendix D</b> | <b>Data mining prediction models of coal and coke quality .....</b>                           | <b>338</b> |
| <b>Appendix E</b> | <b>Sub-models of the cokemaking process .....</b>   | <b>341</b> |
| <b>Appendix F</b> | <b>Collated mineral data of coals .....</b>   | <b>342</b> |
| <b>Appendix G</b> | <b>Extended discussion on vitrinite reflectance distributions .....</b>                       | <b>351</b> |
| G.1               | Preliminary Modelling Results.....  | 351        |
| G.1.1             | Implementation of the SOM on Original Data .....  | 358        |
| <b>Appendix H</b> | <b>Decision tree assessment of vitrinite reflectance distributions .....</b>                  | <b>362</b> |
| H.1               | Decision Tree Analysis of Model Residuals .....   | 362        |

|  |   |            |
|--|---|------------|
| H.1.1  | Original Regression Residuals.....                | 362        |
| H.1.2  | Exploration of V-Group Regression Residuals ..... | 364        |
| H.1.3  | Discussion of residual analysis.....              | 366        |
| <b>Appendix I Assessment of fusibility calculations (Coin and co-workers).....</b> |   | <b>368</b> |
| I.1  | Overview .....                                    | 368        |
| I.1.1  | Point Count of Coal and Coke.....                 | 369        |
| I.1.2  | Determine Maceral Densities .....                 | 370        |
| I.1.3  | Convert Point Count to Weight Basis.....          | 372        |
| I.1.4  | Convert Volatile Matter to Maceral Basis .....    | 372        |
| I.1.5  | Determine Volatile Matter Correction Factor ..... | 373        |
| I.1.6  | Determine Percentage of Maceral in Coke.....      | 373        |
| I.1.7  | Determine Percentage Fusing.....                  | 374        |
| I.1.8  | Uncertainty Calculations .....                    | 374        |
| <b>Appendix J Summary of mineral behaviour .....</b>                               |   | <b>376</b> |
| <b>Appendix K Surface plots of mineral effect and CRI .....</b>                    |   | <b>380</b> |
| <b>Appendix L Summary of models assessed .....</b>                                 |   | <b>383</b> |
| <b>Appendix M Model fit surface plots, by model series .....</b>                   |   | <b>406</b> |
| M.1  | Model Series 1 .....                              | 407        |
| M.1.1  | Model 1a.....                                     | 407        |
| M.1.2  | Model 1b.....                                     | 408        |
| M.1.3  | Model 1c.....                                     | 409        |
| M.1.4  | Model 1d.....                                     | 410        |
| M.1.5  | Model 1e.....                                     | 411        |
| M.1.6  | Model 1f .....                                    | 412        |
| M.2  | Model Series 2 .....                              | 413        |
| M.2.1  | Model 2a .....                                    | 413        |

|       |                     |     |
|-------|---------------------|-----|
| M.2.2 | Model 2b.....       | 418 |
| M.2.3 | Model 2c.....       | 423 |
| M.2.4 | Model 2d.....       | 428 |
| M.2.5 | Model 2e.....       | 433 |
| M.2.6 | Model 2f.....       | 438 |
| M.3   | Model Series 3..... | 443 |
| M.3.1 | Model 3a.....       | 443 |
| M.3.2 | Model 3b.....       | 444 |
| M.3.3 | Model 3c.....       | 445 |
| M.3.4 | Model 3d.....       | 446 |
| M.3.5 | Model 3e.....       | 447 |
| M.3.6 | Model 3f.....       | 448 |
| M.4   | Model Series 4..... | 449 |
| M.4.1 | Model 4a.....       | 449 |
| M.4.3 | Model 4d.....       | 450 |
| M.5   | Model Series 5..... | 451 |
| M.5.1 | Model 5a.....       | 451 |
| M.5.2 | Model 5b.....       | 456 |
| M.5.3 | Model 5c.....       | 461 |
| M.5.4 | Model 5d.....       | 466 |
| M.5.5 | Model 5e.....       | 471 |
| M.5.6 | Model 5f.....       | 476 |
| M.6   | Model Series 6..... | 481 |
| M.6.1 | Model 6a.....       | 481 |
| M.6.2 | Model 6b.....       | 494 |
| M.6.3 | Model 6c.....       | 507 |

|       |                     |     |
|-------|---------------------|-----|
| M.6.4 | Model 6d.....       | 520 |
| M.6.5 | Model 6e.....       | 533 |
| M.6.6 | Model 6f.....       | 546 |
| M.7   | Model Series 7..... | 559 |
| M.7.1 | Model 7a.....       | 559 |
| M.7.2 | Model 7d.....       | 560 |
| M.8   | Model Series 8..... | 561 |
| M.8.1 | Model 8a.....       | 561 |
| M.8.2 | Model 8b.....       | 574 |
| M.8.3 | Model 8c.....       | 587 |
| M.8.4 | Model 8d.....       | 600 |
| M.8.5 | Model 8e.....       | 613 |
| M.8.6 | Model 8f.....       | 626 |

# List of Figures

Figure 2.1: Negative linear correlation between CSR and CRI. Blends are coloured red, single coals are coloured blue, unknown components or both singles and blends are coloured green. Operational results are presented as solid lines, whilst experimental, laboratory, and unreported data sources are shown as dashed line, dash-dot line, and dotted line respectively. ....31

Figure 2.2: Relationship between vitrinite reflectance and the modification of CRI. Note this chart represents the coefficients associated with vitrinite reflectance only and does not include constants or other model parameters. Blends are coloured red, Single coals are coloured blue, unknown components or both singles and blends are coloured green. Operational results are presented as solid lines, whilst experimental, laboratory, and unreported data sources are shown as dashed line, dash-dot line, and dotted line respectively. ....35

Figure 2.3: Relationship between vitrinite reflectance and the modification of CSR. Note this chart represents the coefficients associated with vitrinite reflectance only and does not include constants or other model parameters. Blends are coloured red, Single coals are coloured blue, unknown components or both singles and blends are coloured green. Operational results are presented as solid lines, whilst experimental, laboratory, and unreported data sources are shown as dashed line, dash-dot line, and dotted line respectively. ....36

Figure 2.4: Relationship between volatile matter and the modification of CRI. Note this chart represents the coefficients associated with volatile matter only and does not include constants or other model parameters. Blends are coloured red, Single coals are coloured blue, unknown components or both singles and blends are coloured green. Operational results are presented as solid lines, whilst experimental, laboratory, and unreported data sources are shown as dashed line, dash-dot line, and dotted line respectively. ....38

Figure 2.5: Relationship between volatile matter and the modification of CSR. Note this chart represents the coefficients associated with volatile matter only and does not include constants or other model parameters. Blends are coloured red, Single coals are

coloured blue, unknown components or both singles and blends are coloured green. Operational results are presented as solid lines, whilst experimental, laboratory, and unreported data sources are shown as dashed line, dash-dot line, and dotted line respectively. ....39

Figure 2.6: Relationship between basicity index and the modification of CRI. Note this chart represents the coefficients associated with basicity only and does not include constants or other model parameters. Blends are coloured red, Single coals are coloured blue, unknown components or both singles and blends are coloured green. Operational results are presented as solid lines, whilst experimental, laboratory, and unreported data sources are shown as dashed line, dash-dot line, and dotted line respectively. ....42

Figure 2.7: Relationship between basicity index and the modification of CSR. Note this chart represents the coefficients associated with basicity only and does not include constants or other model parameters. Blends are coloured red, Single coals are coloured blue, unknown components or both singles and blends are coloured green. Operational results are presented as solid lines, whilst experimental, laboratory, and unreported data sources are shown as dashed line, dash-dot line, and dotted line respectively. ....43

Figure 2.8: Relationship between fluidity and the modification of CSR and CRI. Note this chart represents the coefficients associated with fluidity only and does not include constants or other model parameters. Blends are coloured red, Single coals are coloured blue, unknown components or both singles and blends are coloured green. Operational results are presented as solid lines, whilst experimental, laboratory, and unreported data sources are shown as dashed line, dash-dot line, and dotted line respectively. ....50

Figure 2.9: Relationship between Audibert-Arnu Dilatation and the modification of CRI. Note this chart represents the coefficients associated with dilatation only and does not include constants or other model parameters. Blends are coloured red, Single coals are coloured blue, unknown components or both singles and blends are coloured green. Operational results are presented as solid lines, whilst experimental, laboratory, and unreported data sources are shown as dashed line, dash-dot line, and dotted line respectively. ....53

Figure 2.10: Relationship between Audibert-Arnu Dilatation and the modification of CSR. Note this chart represents the coefficients associated with dilatation only and does not include constants or other model parameters. Blends are coloured red, Single coals

|   |     |
|---|-----|
| are coloured blue, unknown components or both singles and blends are coloured green. Operational results are presented as solid lines, whilst experimental, laboratory, and unreported data sources are shown as dashed line, dash-dot line, and dotted line respectively. ....   | 54  |
| Figure 2.11: Relationship between Sapozhnikov Plastic Layer Thickness and the modification of CSR and CRI. Note this chart represents the coefficients associated with layer thickness only and does not include constants or other model parameters. Blends are coloured red, Single coals are coloured blue, unknown components or both singles and blends are coloured green. Operational results are presented as solid lines, whilst experimental, laboratory, and unreported data sources are shown as dashed line, dash-dot line, and dotted line respectively. .... | 55  |
| Figure 3.1: Flow diagram of the CRISP-DM Process. ....  | 79  |
| Figure 3.2: Artificial Neural Network Map, showing an input layer with three nodes, a single hidden layer with four nodes, and an output layer with two nodes. ....   | 106 |
| Figure 5.1: Left - Vitrinite reflectance distribution of a blend containing two coals (coloured as different shades of green) of varying vitrinite reflectance. The percentage of points counted for a sample at a given reflectance is shown. The average vitrinite reflectance of the two coals is shown as the dotted line, whilst the grey box indicates a sum across a subset of the reflectance range. Right – two vitrinite reflectance distributions with the same average vitrinite reflectance and standard deviation. ....                                       | 127 |
| Figure 5.2: Offsetting distributions to a central reflectance value. ....   | 130 |
| Figure 5.3: VDCs derived from the self-organizing map. ....   | 133 |
| Figure 5.4: Original regression fits, grouped by VDCs, and coloured by percentage Australian coals, where the darker the data point, the higher the proportion of Australian coal. ....   | 135 |
| Figure 5.5: Original regression blend fit and overall model statistics for both single coals and blends. ....   | 136 |
| Figure 5.6: V-Group regression blend fit and overall model statistics for both single coals and blends. ....  | 136 |

|  |     |
|--|-----|
| Figure 5.7: VDCs sorted by model coefficient (smallest at left, greatest at right). The greater the model coefficient, the greater the underprediction by the original model. ....   | 137 |
| Figure 5.8: Flowchart describing the extraction of rules. ....   | 140 |
| Figure 5.9: Misclassified instances from the decision tree, grouped by SOM actual VDC, coloured by class assigned by decision tree. ....   | 150 |
| Figure 6.1: Ascending calculated fully or partially fused semifusinite. Primary vertical axis is the proportion of calculated fully or partially fusing semifusinite, secondary vertical axis is the number of samples measured for each coal. ....  | 161 |
| Figure 6.2: Ascending fully fusing semifusinite for each coal. Primary vertical axis is the proportion of calculated fully fusing semifusinite, secondary vertical axis is the number of samples measured for each coal. ....  | 161 |
| Figure 6.3: Frequency of coals in the data set. ....   | 162 |
| Figure 6.4: J48 decision tree to identify key distinguishing features of coal Class C. Attributes shown in the decision tree are total sulfur (TotSul), micrinite maceral (Mic), vitrinite maceral group (Vit), ultimate nitrogen (UltNdaf), potassium oxide (K2O), macrinite maceral (Mac), and thermoplastic behaviour (Dil_C). Interpretation of the tree structure - white ovals are decision nodes, and grey boxes are terminating nodes, which detail the class assigned to that node (Other Class or Class C coal), and the number of instances assigned to that node / number of instances incorrectly classified to that node. Figure visualised in GraphViz 2.38 (Ellson et al. 2001)..... | 169 |
| Figure 6.5: Scree plot of model performance as function of number of clusters.....   | 171 |
| Figure 6.6: Three dimensional plot of model performance. Training performance is shown by the surface gradient, number of clusters is shown at nodes marked by small coloured squares. The dark contour shown inside the volume represents a test performance of 0.765. Results visualised in Paraview 5.4.1 (Ahrens et al. 2005; Ayachit 2015). ....  | 173 |
| Figure 6.7: Examination of SMOTE rate and number of nearest neighbours at a cluster influence range of 0.35. The optimum value circled yields the best training and test model fits, whilst minimising SMOTE rate and number of clusters. Results visualised in Paraview 5.4.1 (Ahrens et al. 2005; Ayachit 2015).....   | 173 |



|  |     |
|--|-----|
| Figure 6.8: Model performance, predicted versus measured values for the prediction of coke inerts, grouped by coal category. A category beginning with ‘s’ indicates synthetic data. Lines represent the $x=y$ lines, and the 95% confidence interval on the measured values, as determined in Section 6.5.4. .... | 174 |
| Figure 6.9: Surface plots of variable behaviour. in1 to in6 are $Ash_{db}$ , $UltC_{daf}$ , $UltS_{daf}$ , $H_{excess}$ , Micrinite, and $P_{coal}$ respectively, and out1 is the predicted coke inerts. ....  | 176 |
| Figure 6.10: Relationship between model performance and SMOTE rate. Increasing the SMOTE rate improves both $r^2$ training and testing values rapidly up to the rate of 400%. ....   | 178 |
| Figure 7.1: Model schematic of proposed back-calculation of mineral forms. ....  | 192 |
| Figure 7.2: Predicted mineral composition plotted against the measured composition. The error bars represent the range of the 10 fold repetition, whilst the solid diagonal line represents $x=y$ . The green dashed line is the linear fit between the measured and fitted results.....                           | 197 |
| Figure 7.3: Histograms of predicted mineral composition, as percentage in coal. ....   | 203 |
| Figure 7.4: Correlation matrix using Spearman-Rho Correlation. ....  | 206 |
| Figure 7.5: Scree plot of explained variance from PCA results. ....  | 207 |
| Figure 7.6: Biplot of the first two PCA components. ....   | 207 |
| Figure 7.7: Results of Boruta feature selection, showing the significance of features above the shadow (random representations) of the features.....   | 209 |
| Figure 7.8: RMSE for RFE Feature Selection. Error bars show the standard deviation, whilst the point marked in blue (feature 13) is representative of the last included feature. ....  | 210 |
| Figure 7.9: Graphical representation of support vector machines. a) A linearly separable case. b) A non-linearly separable case. c) Transformation of non-linear data into a higher dimension using the kernel trick.....  | 212 |
| Figure 7.10: Fitted v Measured values for the linear regression model and the support vector machine. Lines represent fitted=measured, and measured $\pm 5$ .....  | 216 |

|  |     |
|--|-----|
| Figure 7.11: Fitted v Measured values for the linear regression model and the support vector machine with the additional chamosite term. Lines represent fitted=measured, and measured +/- 5.....  | 219 |
| Figure 7.12: Surface plot of the relationship between CRI, volatile matter, and modified basicity index. ....  | 220 |
| Figure 7.13: Fitted v Measured values for the linear regression model and the support vector machine using only the modified basicity index and volatile matter as input terms. Lines represent fitted=measured, and measured +/- 5. ....  | 221 |
| Figure 8.1: Correlation Matrix. ....   | 232 |
| Figure 8.2: Behaviour of VDCs (as dtClust) with respect to CSR, with varying fluidity (top) and volatile matter (bottom). Average values were used for the remaining inputs. ....  | 240 |
| Figure 8.3: VDCs sorted by Chapter 5 model coefficient (smallest at left, greatest at right). Figure replicated from Chapter 5. ....   | 241 |
| Figure 8.4: VDCs sorted by fluidity offset from the SVM prediction (Model 2e). The highest optimum fluidity is shown at left, whilst the lowest optimum fluidity is at the right. ....   | 241 |
| Figure 8.5: Behaviour of selected minerals with respect to CSR for Model 5e. Mineral percentage reported on a percentage in coal basis. Average values were used for the remaining inputs. Note that the x-axis range presented here extends beyond the predicted range for some of the minerals which is represented by the point at which plotted behaviour becomes a constant CSR value. .... | 246 |
| Figure 8.6: Performance of model 5d which includes attributes VMdaf, Dil_CD, and selected minerals. Top, left: overall fitted v measured, Top, right: q-q plot. Bottom, left: residual v fits, Bottom, right: blend fitted v measured. Blue colour represents the SVM model points, and the black lines on fitted plots show a 1:1 fit line, with lines +/- 5 units.....                         | 253 |
| Figure 8.7: Top – Bottom, Left – Right. Surface plots of CSR (z axis), with VM <sub>daf</sub> (right y axis) plotted against CaRich, Chamosite, Dil_CD, FeRich, Fluorapatite, Kaolinite, and Quartz, respectively.....   | 254 |

Figure 8.8: Change in predicted CSR in response to varying the volatile matter. Model fits have been forced through zero. Where there are interactions with other parameters, constants have been substituted, as per Appendix A. The inset image shows a limited region of CSR, where the results of the present study fall (red curved line in centre of figure).....255

# List of Tables

|  |     |
|--|-----|
| Table 1.1: Proximate analysis measures and associated ISO Standards. Descriptions make reference to both the standards, and comments provided in Speight (2015) and Zhu (2014). .....  | 8   |
| Table 1.2: Ultimate analysis measures and associated ISO Standards. Descriptions make reference to both the standards, and comments provided in Speight (2015) and Zhu (2014). .....   | 9   |
| Table 1.3: Maceral Groups. Reproduced from Speight (2012). .....   | 10  |
| Table 1.4: Measures of caking properties and thermoplastic behaviour, and associated ISO Standards. Descriptions make reference to both the standards, and comments provided in Zhu (2014) and Thomas and Thomas (2002). ..... | 12  |
| Table 2.1: Comparison of definitions of optical textures. ....   | 58  |
| Table 2.2: Comparison of relationships between coal quality parameters and Micum Indices. ....   | 72  |
| Table 3.1: Data Type and Data Source as Reported by Authors. Correlation coefficients reported in grey boxes for regressions for CSR and CRI only ( $N_{\text{Models}} = 110$ ). ....  | 82  |
| Table 3.2: Comparison of controlled parameters. Grey fill indicates operational data. ....   | 87  |
| Table 3.3: Coal quality controls and ranges. Grey fill indicates operational data. ....  | 90  |
| Table 3.4: Numbers of variables in models for CSR or CRI prediction (excluding constants). ....  | 102 |
| Table 3.5: Neural networks used, and their associated distinguishing features.....   | 108 |
| Table 3.6: Training algorithms used, and their distinguishing features.....  | 109 |
| Table 3.7: Activation functions applied, and their characteristics.....  | 111 |
| Table 3.8: Reported training and test set splits.....  | 113 |
| Table 5.1: Correlation matrix of created attributes.....   | 142 |
| Table 5.2: Sample confusion matrix. ....   | 144 |

|  |     |
|--|-----|
| Table 5.3: Confusion Matrix associated with CART decision tree. ....   | 145 |
| Table 5.4: CART decision tree performance metrics. ....  | 146 |
| Table 5.5: Confusion Matrix associated with C4.5 decision tree. ....   | 147 |
| Table 5.6: C4.5 decision tree performance metrics. ....  | 147 |
| Table 5.7: Parameter optimisation table. ....  | 149 |
| Table 5.8: Confusion Matrix associated with optimised C4.5 decision tree. ....   | 149 |
| Table 6.1: Comparison of model fit including a measure for fusibility.....   | 180 |
| Table 7.1: Minerals identified in coal. List from Longbottom and Monaghan (2015),<br>coloured by their abundance in coals, as reported in Reid and Monaghan (2012) for US<br>coal basins, and by Grigore (2007) for Australian coals (using cut offs of <1.5% ash<br>weight basis for “rare”, >5% as “common”, and those not meeting the criteria but<br>identified as “variable”). Where conflicting classifications are identified, those for<br>Australian coals have been applied..... | 187 |
| Table 7.2: Assumed form of minerals where variable composition is an issue.....  | 190 |
| Table 7.3: Selected minerals for analysis .....  | 191 |
| Table 7.4: Comparison of model performance on Ward’s results for carbonate altered<br>lithic sandstone (From Cohen and Ward (1991)). .....   | 195 |
| Table 7.5: Individual mineral basicity values. ....  | 198 |
| Table 7.6: Filtering cut offs for minerals. ....   | 204 |
| Table 7.7: Range of parameters considered. ....  | 215 |
| Table 8.1: Parameter filtering ranges for the final model series.....  | 228 |
| Table 8.2: Example of one-hot encoding. ....   | 230 |
| Table 8.3: Comparison of model behaviour associated with rank and thermoplastic<br>terms.....  | 235 |
| Table 8.4: ANOVA Analysis comparing VDC models (Series 2) to Base Case (Series<br>1) models. Where the Base model reported better fit, - is reported. N/A is reported in<br>the case where the degrees of freedom is too low to report significance. Those that are<br>not statistically significant at the 5% level are shaded grey. ....   | 238 |

|   |     |
|---|-----|
| Table 8.5: ANOVA Analysis comparing cokeInerts models (Series 3) to Base Case (Series 1) models. Where the Base model reported better fit, - is reported. N/A is reported in the case where the degrees of freedom is too low to report significance. Those that are not statistically significant at the 5% level are shaded grey.....   | 243 |
| Table 8.6: ANOVA Analysis comparing coal minerals models (series 5) to Base Case (series 1) models. Where the Base model reported better fit, - is reported. N/A is reported in the case where the degrees of freedom is too low to report significance. Those that are not statistically significant at the 5% level are shaded grey.....  | 245 |
| Table 8.7: ANOVA Analysis comparing coal minerals and VDC models (Series 8) to coal minerals (Series 5) models. Where the Base model reported better fit, - is reported. N/A is reported in the case where the degrees of freedom is too low to report significance. Those that are not statistically significant at the 5% level are shaded grey. Cases where Series 5 was not significant are marked as “S5 N.S.” ..... | 248 |
| Table 8.8: Averaged model performance (across the different rank and thermoplastic combinations) by series. * Indicates that only two models are presented within the series instead of six, meaning the average reported in the table may not be representative. Bold text shows the best averaged model series fit. ....  | 249 |
| Table 8.9: Averaged blend performance (across the different rank and thermoplastic combinations) by series. * Indicates that only two models are presented within the series instead of six, meaning the average reported in the table may not be representative. Bold text shows the best averaged model series fit. ....  | 251 |