Applications of Process Analytics and Machine Learning in Pyrometallurgy and Kraft Pulping

by

Lee Dale Rippon

BASc, The University of British Columbia, 2014

MASc, The University of British Columbia, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Chemical and Biological Engineering)

The University of British Columbia
(Vancouver)

September 2023

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Applications of Process Analytics and Machine Learning in Pyrometallurgy and Kraft Pulping

submitted by **Lee Dale Rippon** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Chemical and Biological Engineering**.

Examining Committee:

Bhushan Gopaluni, Professor, Chemical and Biological Engineering, UBC
Supervisor

Philip D. Loewen, Professor, Mathematics, UBC
Co-Supervisor

Susan A. Baldwin, Professor, Chemical and Biological Engineering, UBC
University Examiner

Chunping Dai, Associate Professor, Wood Science, UBC
University Examiner

Additional Supervisory Committee Members:

Paul Bicho, Manager of Innovation and Optimization at Canfor Pulp
Supervisory Committee Member

Yankai Cao, Assistant Professor, Chemical and Biological Engineering, UBC
Supervisory Committee Member

Devin Marshman, Solutions Architect at Spartan Controls
Supervisory Committee Member

Abstract

The optimization of legacy industrial processes is critical for the economic viability of many rural communities in Canada. Automation and advanced process control is of paramount importance for many large-scale industrial processes to maintain viability in a constricting regulatory environment that is increasingly competitive economically. However, with legacy industrial processes comes a rich history of automation, including large quantities of underappreciated process data. This dissertation is about leveraging existing historical data with machine learning and process analytics to generate novel data-driven solutions to outstanding process faults. An application-driven approach provides insights into the full stack of considerations including identifying and framing data-driven opportunities for control of complex industrial processes, acquiring the necessary resources, preparing the data, developing and evaluating methods, and deploying sustainable solutions. Contributions are made to help address highly troublesome faults in two distinct industrial processes.

The first industrial case study involves mitigating the impact of unexpected loss of plasma arc in an electric arc furnace that is key to a 60,000 tonne/year pyrometallurgy operation. A convolutional neural network classifier is trained to learn a representation from the operating data that enables prediction of the arc loss events. The operating data and problem formulation are published as a novel benchmark challenge to address observed shortcomings with existing fault detection benchmark literature.

The second industrial case study involves advanced monitoring of a rotary lime kiln

in a 152,000 tonne/year kraft pulp mill to mitigate faults such as ring formation and refractory wear. A novel shell temperature visualization strategy is published that enables improved monitoring and empowers researchers and industry professionals to obtain value from thermal camera data. Various approaches are studied for monitoring ring formation. Aberrations in shell temperatures led to the discovery of a novel phenomenon known as rotational aliasing that has important implications for measurement and analysis of shell temperature data. Finally, inferential sensing of residual calcium carbonate content is studied to help optimize specific energy and reduce emissions.

Lay Summary

In recent decades, a powerful secular trend towards digital transformation has emerged in advanced manufacturing, and society at large. Industrial chemical and biological processes are deploying advanced measurement and instrumentation technologies that produce an increasingly overwhelming amount of data. This has triggered a torrent of research and investment in data-driven process optimization and advanced process control. However, these capital intensive and safety sensitive operating environments are also inherently risk averse. This research aims to help bridge the divide between the ambitions of artificial intelligence and the assurances required by existing industrial operations. Raw data from large scale industrial processes is used to derive operating insights and engineer tools for advanced process control. Specifically, operating data from a 60,000 tonne/year pyrometallurgical plant and a 152,000 tonne/year kraft pulp mill is combined with process knowledge and machine learning to develop algorithms that can improve safety, reduce environmental impacts, and increase profitability.

Preface

During the years of research that went into preparing this dissertation there were two distinct collaborations with industry partners along with various collaborations with peers in academia. All of the work presented was conducted in the Data Analytics and Intelligent Systems Laboratory at the University of British Columbia.

The first two chapters discuss assorted machine learning and process analytics applications for advanced process control in collaboration with, Prof. Bhushan Gopaluni, Prof. Philip Loewen, Dr. Michael Forbes, Johan Backstrom, Prof. Qiugang Lu, Dr. Yiting Tsai, Dr. Kai Wang, Siang Lim, and Shams Elnawawi with the following research output:

- Q. Lu, L. D. Rippon, R. B. Gopaluni, M. G. Forbes, P. D. Loewen, J. Backström, and G. A. Dumont. Noncausal modeling and closed-loop optimal input design for cross-directional processes of paper machines. In *2017 American Control Conference (ACC)*, pages 2837–2842. IEEE, 2017

- Y. Tsai, Q. Lu, L. Rippon, S. Lim, A. Tulsyan, and B. Gopaluni. Pattern and knowledge extraction using process data analytics: A tutorial. *IFAC-PapersOnLine*, 51(18):13–18, 2018

- L. D. Rippon, Q. Lu, M. G. Forbes, R. B. Gopaluni, P. D. Loewen, and J. U. Backström. Machine direction adaptive control on a paper machine. *Industrial & Engineering Chemistry Research*, 58(26):11452–11473, 2019

- K. Wang, L. Rippon, J. Chen, Z. Song, and R. B. Gopaluni. Data-driven dy-

namic modeling and online monitoring for multiphase and multimode batch processes with uneven batch durations. *Industrial & Engineering Chemistry Research*, 58(30):13628–13641, 2019

- S. C. Lim, S. Elnawawi, L. D. Rippon, D. L. O'Connor, and R. B. Gopaluni. Data quality over quantity: Pitfalls and guidelines for process analytics. *IFAC World Congress 2023*, pages 1–8, 2023

I was the lead investigator for the 2019 publication on machine direction adaptive control. I was responsible for concept formation, data analysis, and manuscript composition whereas the co-authors supported concept formation and manuscript review. For the remaining publications the first author was the lead investigator while myself and the remaining authors had supporting responsibilities on data analysis, concept improvement, and manuscript composition.

Chapter 3 presents research on fault detection in a pyrometallurgy process in collaboration with BBA Engineering Consultants. Contributions were made by Prof. Bhushan Gopaluni, Prof. Sirish Shah, Dr. Carole Prévost, Michel Ruel, and Ibrahim Yousef with the following research output:

- Foundations of Process Analytics and Machine Learning (FOPAM) 2019, Raleigh NC. Poster presentation on "Representation Learning for Inferential Sensor Development in an Electric Arc Furnace."

- L. D. Rippon, I. Yousef, R. B. Gopaluni, B. Hosseini, J. F. Beaulieu, C. Prévost, and S. L. Shah. Process analytics and machine learning to predict arc loss in an electric arc furnace. In *59th Conference of Metallurgists 2020 hosting the 4th International Uranium Conference*, 2020.

- The 3rd BC Universities Systems and Control Meeting, Victoria BC, Aug. 2020. Presentation on "Process Analytics and Machine Learning to Predict Arc Loss in an Electric Arc Furnace." Awarded Best Presentation Award.

- Emerging Technologies: BC's AI Showcase Conference by UBC's Centre for Artificial Intelligence Decision-making and Action (CAIDA), Vancouver BC, Nov. 2020. Poster presentation on "Process Analytics and Machine

Learning to Predict Arc Loss in an Electric Arc Furnace."

- I. Yousef, R. B. Gopaluni, L. D. Rippon, S. L. Shah, J. F. Beaulieu, and C. Prévost. Poster presentation on "Process Data Analytics and Representation Learning to Predict Arc Loss in an Electric Furnace" at 2020 Virtual AIChE Annual Meeting .

- L. D. Rippon, I. Yousef, B. Hosseini, A. Bouchoucha, J. F. Beaulieu, C. Prévost, M. Ruel, S. Shah, and R. B. Gopaluni. Representation learning and predictive classification: Application with an electric arc furnace. *Computers & Chemical Engineering*, 150:107304, 2021

- I. Yousef, L. D. Rippon, C. Prévost, S. L. Shah, and R. B. Gopaluni. The arc loss challenge: A novel industrial benchmark for process analytics and machine learning. *Journal of Process Control*, 128:103023, 2023

For this industrial case study I was initially lead investigator, then I. Yousef and myself were co-lead investigators, and towards the end of this work I. Yousef led and I supported. For all of these publications I. Yousef and myself closely shared in the majority of responsibilities (e.g., data analysis and concept formation), with first authorship indicating who led manuscript composition. The presentations listed here were delivered by me. The remaining co-authors provided support with data acquisition, conceptual guidance, and manuscript review.

Chapter 4 presents research on fault detection and soft sensor development for rotary lime sludge kilns in collaboration with Canfor Pulp and Spartan Controls. Contributions were made by Prof. Bhushan Gopaluni, Prof. Philip Loewen, Dr. Peter Gorog, Devin Marshman, Carl Sheehan, Dr. Paul Bicho, Barry Hirtz, Travis Reinheimer, and Cilius van der Merwe with the following research output:

- PACWEST 2021 Technical Conference. Presentation on "Detection and Diagnosis of Ring Formation in Rotary Lime Kilns - Part I Developing a Ring Formation Indicator." Awarded Best Student Paper Award [104].

- L. Rippon, B. Hirtz, C. Sheehan, T. Reinheimer, P. Loewen, and B. Gopaluni. Visualization of multiscale ring formation in a rotary kiln. *Nordic Pulp &*

Paper Research Journal, 36(4):549–558, 2021

- L. D. Rippon, B. Hirtz, C. Sheehan, T. Reinheimer, C. van der Merwe, P. Loewen, and B. Gopaluni. Detection and diagnosis of ring formation in rotary lime kilns. *Canadian Chemical Engineering Conference Proceedings*, pages 23–29, 2021

- L. D. Rippon, B. Hirtz, C. Sheehan, T. Reinheimer, C. van der Merwe, P. Loewen, and B. Gopaluni. Rotary kiln monitoring with shell temperature visualization and process analytics. In *2022 TAPPI PEERS and IBBC Conference Proceedings*. TAPPI Press, 2022

I was lead investigator for this industrial case study and I was responsible for leading the data analysis, concept formation, and manuscript composition for all of the items listed above. The co-authors provided support with important feedback on the concepts and methods of analysis, help with data acquisition, and assistance with manuscript review.

Table of Contents

Abstract . . . . . **iii**

Lay Summary . . . . . **v**

Preface . . . . . **vi**

Table of Contents . . . . . **x**

List of Tables . . . . . **xiii**

List of Figures . . . . . **xiv**

List of Abbreviations . . . . . **xx**

Acknowledgments . . . . . **xxiii**

1 Introduction **1**

1.1 Machine Learning and Process Analytics 1

1.2 Motivation 3

1.3 Objectives 5

1.3.1 Improving operating outcomes with industrial data 6

1.3.2 Predicting arc loss in an electric arc furnace 6

1.3.3 Lime kiln monitoring with infrared thermal cameras . . . 7

1.4 Thesis Outline and Contributions 7

2 Industrial Applications of Process Analytics and Machine Learning 10

2.1 Fault Detection and Diagnosis in Chemical and Biological Processes 11
2.1.1 Process history based methods 12
2.1.2 Performance benchmarking - limitations and opportunities 14
2.1.3 Deep learning for fault detection 16
2.2 Learning from Industrial Process Data 21
2.2.1 Improving operating outcomes with process analytics . . . 21
2.2.2 Challenges while learning from industrial process data . . 27

3 Predicting Arc Loss in an Electric Arc Furnace 31
3.1 Introduction to Arc Loss in an Electric Furnace 33
3.2 Data Preparation and Visualization 36
3.2.1 The arc loss dataset 37
3.2.2 Data structuring and output labeling 37
3.2.3 Data visualization 40
3.2.4 Data cleaning 41
3.3 Learned Representations and Predictive Classifiers 44
3.3.1 Explicit representations with reduced dimensionality . . . 46
3.3.2 Predictive classification and implicit representations . . . 47
3.4 Experimental Setup 49
3.4.1 Stratified k-fold cross-validation 49
3.4.2 Hyperparameter optimization 51
3.5 Results and Discussion of Arc Loss Prediction 54
3.5.1 Experimental results 54
3.5.2 Discussion of contributions and findings 58
3.6 The Arc Loss Challenge 59

4 Lime Kiln Monitoring with Infrared Thermal Cameras 62
4.1 Introduction to Rotary Lime Sludge Kilns 64
4.2 Thermal Cameras and Data Visualization 66
4.2.1 Applications of thermal cameras on lime kilns 67
4.2.2 Kiln shell temperature visualization 69
4.3 Monitoring Ring Formation 79
4.3.1 Ring formation in lime sludge kilns 79

4.3.2 Developing a ring formation indicator 82
4.3.3 Rotational aliasing 89
4.4 Inferential Sensing for Residual Carbonate Prediction 93
4.4.1 Residual carbonate control for kiln optimization 94
4.4.2 Hyper-inferential sensing 96
4.4.3 Experimental results 102
4.4.4 Discussion and recommendations 104

5 Pitfalls and Guidelines for Industrial Process Analytics 107
5.1 Identifying and Framing a Data-driven Opportunity 107
5.2 Data Acquisition and Contextualization 110
5.3 Data Preparation 113
5.4 Method Selection, Development, and Evaluation 117
5.5 Deploying Sustainable Solutions 119

6 Conclusion 122

Bibliography 124

A Supporting Materials 139
A.1 Pyrometallurgy Case Study 139
A.2 Kraft Pulping Case Study 146

List of Tables

Table 3.1 Hyperparameter selection for traditional machine learning (ML) algorithms. 53
Table 3.2 Hyperparameter selection for deep learning algorithms. 54
Table 3.3 Summary of the experimental results. 55

Table 4.1 Validating growth and decay indications for precision. 88
Table 4.2 Summary of results from testing residual carbonate predictions developed with the proposed hyper-inferential sensing framework. 104

Table A.1 Overview of operating variables analyzed to predict arc loss for the pyrometallurgy case study. 139
Table A.2 Hyperparameter search space for traditional ML algorithms. . . 144
Table A.3 Hyperparameter search space for deep learning algorithms. . . 144
Table A.4 Supplemental experimental result metrics. 145
Table A.5 Overview of operating variables analyzed to investigate ring formation and predict residual calcium carbonate for the kraft pulping case study. 146
Table A.6 Hyperparameter search space for traditional ML algorithms. . . 149
Table A.7 Hyperparameter search space for deep learning algorithms. . . 149

List of Figures

Figure 1.1 The field of machine learning includes three pillars, i.e., supervised learning, unsupervised learning, and reinforcement learning. 2

Figure 2.1 Stages involved with delivering a novel data-driven solution. . 15

Figure 2.2 A summary of deep learning techniques for fault detection and diagnosis (FDD) on Tennessee Eastman process (TEP) benchmark dataset including feedforward artificial neural networks (ANNs) [43], generative adversarial networks (GANs) [34, 52, 74, 120, 145], autoencoders (AEs) [21, 51, 81, 82, 144, 150, 154], convolutional neural networks (CNNs) [18, 20, 35, 101, 142], and recurrent neural networks (RNNs) [19, 123, 152, 153, 155]. 19

Figure 2.3 Illustration of the CNN predictive classifier architecture [108]. 20

Figure 2.4 Lost production due to upsets of varying duration for the furnace feed, furnace power, and calciner feed in a pyrometallurgical process. 23

Figure 2.5 Lost production due to upsets of varying duration for the mud filter feeds, kiln feed, and kiln fuel usage in the recausticizing area of a kraft pulp mill. 24

Figure 3.1 Simplified illustration of the relevant mining and metallurgical processes [106]. 34

Figure 3.2 An illustration of a direct current electric arc furnace [106]. . . 35

Figure 3.3 Flowchart illustrating the overall data analytics workflow. . . . 36
Figure 3.4 Structuring consecutive days of historical data. 38
Figure 3.5 Illustration and quantitative definition of conditions constituting arc loss. 39
Figure 3.6 Visual validation of the arc loss labels. 40
Figure 3.7 Daily arc loss events in each electrode over one year of operation. 41
Figure 3.8 Visually comparing relatively stable operation (top) to faulty operation (bottom) to observe the impact of arc loss upsets [108]. 42
Figure 3.9 Setting process variable (PV) limits with process knowledge to remove invalid data. 43
Figure 3.10 Illustration of data segmentation to create a balanced dataset. . 45
Figure 3.11 Experimental configurations with representations (left) and predictive classifiers (right). 46
Figure 3.12 Splitting the data into training and testing sets based on date. . 50
Figure 3.13 Each experimental configuration is trained and validated with stratified k-fold cross-validation [108]. 50
Figure 3.14 Selecting the number of components for principal component analysis (PCA) and partial least squares (PLS) representations with cross-validation. Note that a higher accuracy is obtained with the supervised PLS algorithm. 52
Figure 3.15 Comparing the classification accuracy of logistic regression with convolutional neural networks while varying the amount of training data. 57

Figure 4.1 Contextualizing the lime kiln (outlined in red) within the integrated forest products industry 64
Figure 4.2 A simplified illustration of a rotary lime kiln with three thermal cameras. Lime mud is fed to the kiln where it is dried into a powder and then agglomerated into nodules which are calcined into lime product. 65

Figure 4.3 A stationary two-dimensional infrared (IR) thermal camera measures shell temperatures along a rotary lime kiln [103]. The thermal camera can collect shell temperatures along the diagonal profile line or average the temperatures within the rectangular measurement areas. 67
Figure 4.4 Thermal cameras can be used to monitor for kiln fouling and ring formation. **Left:** fouling in the chain section of the kiln from soda balls. **Right:** two distinct cases of ring formation [107]. 68
Figure 4.5 Synthetic data illustrating a low-resolution map of kiln shell temperature (KST) profiles. 70
Figure 4.6 Thermal camera measurements are averaged over a day and stitched together to generate a KST profile. Measurements from a clean kiln on 2019-01-01 are compared with measurements from 2019-01-20 to emphasize potential ring formation. . . . 71
Figure 4.7 Synthetic illustration of a shell scan measured over a single rotation period. These visualizations are produced by one-dimensional line scanners that use a rotating sensor calibrated to the kiln rotation speed. 71
Figure 4.8 A KST profile with non-equispaced measurement positions. Discrete measurement areas summarize thermal camera pixels into an average temperature at fixed positions and intervals. . . 73
Figure 4.9 Piecewise linear interpolation of the KST profile. The bottom color bar embeds the y-axis temperatures with a uniform, sequential colormap enabling a clear and intuitive single-axis visualization. 73
Figure 4.10 Spatiotemporal heatmap of KST profiles spanning over 5 years of operation. This is the first published visualization that offers a clear and intuitive view of shell temperature evolution over years of rotary kiln operation. 74

Figure 4.11 Focusing on a period of suspected fouling. Potential ring formation is observed after 2018-04-29 at approximately the 18 m and 45 m positions. Between 2018-05-18 and 2018-06-05 the fouling appears to bridge these separate formations and extend into the 23-40 m range. 76
Figure 4.12 Investigating thermal camera images to confirm suspected ring formation approximately 18 m from the firing end of the kiln just after 2018-04-29. **Left:** an image from the firing end camera taken on 2018-04-29. **Right:** an image taken on 2018-05-06 with cooler shell temperatures in measurement areas E, F, and G. 77
Figure 4.13 Observing mid-kiln camera images to confirm ring formation in the 23-40 m range between 2018-05-18 and 2018-06-05. **Left:** image taken on 2018-05-23. **Right:** image from 2018-05-30 that exhibits cooling across all measurement areas except area A. 77
Figure 4.14 Further reducing the heatmap window and plotting relevant PVs for operating context. **Top:** the spatiotemporal heatmap of KST profiles. **Bottom:** the solids percentage of the lime mud fed to the kiln. 78
Figure 4.15 Ring formation in rotary lime sludge kilns. **Top:** severe ringing restricting the orifice of the kiln in the mid-zone and firing end. **Bottom left:** a mill worker studies ring formation inside a kiln. **Bottom right:** stress fracturing of ring into debris with thermal cycling. 80
Figure 4.16 Statistical forecasting of shell temperatures from a thermal camera at a specific axial position along the kiln [107]. 85
Figure 4.17 Obstruction of measurement areas G, H, and I in the bottom image can corrupt the thermal camera data. 85
Figure 4.18 Daily counts for growth and decay indicators based on shell temperature residuals. 86
Figure 4.19 Validating ring formation indications with raw thermal camera images (top) and forecast results (bottom). 88

Figure 4.20 High frequency variations in shell temperatures measured 24.4 m from the firing end of the kiln. Spurious oscillations corrupt the KST profile and hinder efforts towards ring detection. . . . 90
Figure 4.21 Spatiotemporal heatmap from the thermal camera on the firing end of the kiln providing improved spatial and temporal resolution. 91
Figure 4.22 The correlation coefficient between KST data from a rectangular measurement area 18.3 m from the firing end of the kiln and KST data from various profile line positions denoted by their x-coordinate along the profile line. 92
Figure 4.23 The profile line data with the highest correlation (blue) is plotted with the measurement area data (red) to demonstrate the improved noise reduction. 93
Figure 4.24 Clustering with uniform manifold approximation and projection (UMAP) to visualize the samples of high dimensional industrial process data that are used to train the inferential sensor models (blue). Outlier samples (red) are discarded for modeling but they can be investigated to yield operating insights. . . 99
Figure 4.25 Experimental setup for hyper-inferential sensing. A variety of traditional and contemporary methods for process analytics and ML are selected to provide a highly structured framework for efficient and reliable quantitative comparisons. 100
Figure 4.26 Average validation error of the feature learning methods across 2000 experimental trials. The proportion of trials dedicated to each method is depicted in the pie chart 103
Figure 4.27 Average validation error of the regression methods across 2000 experimental trials. The proportion of trials for each method is displayed in the pie chart 104
Figure 4.28 Testing the candidate residual carbonate soft sensors developed with IR camera data (red) and without IR camera data (green). 105

Figure 5.1 Supplementary sources of information are necessary to contextualize process data and obtain meaningful insights [76]. . 112
Figure 5.2 Histogram of process data with different modes of operation. . 116
Figure 5.3 Visualizing data from different modes of operation. 116

List of Abbreviations

AE autoencoder

AEM abnormal event management

AI artificial intelligence

ANN artificial neural network

APC advanced process control

AR auto-regressive

CL convolutional layer

CNN convolutional neural network

CPR continuous precoat renewal

CSV comma separated value

DC direct current

DCS distributed control system

DNN deep neural network

DRL deep reinforcement learning

EAF electric arc furnace

EDA exploratory data analysis

FCL fully connected layer

FDD fault detection and diagnosis

FDR fault detection rate

FIR finite impulse response

FN false negative

FP false positive

GAN generative adversarial network

GRU gated recurrent unit

ID induced draft

ICA independent component analysis

IQR inter-quartile range

ISA International Society of Automation

IR infrared

KPI key performance indicator

*k***NN** *k*-nearest neighbors

KST kiln shell temperature

K-SVC kernel support vector classifier

KPI key performance indicator

LR logistic regression

LSTM long short-term memory

L-SVC linear support vector classifier

MA moving-average

MAD mean absolute deviation

MAE mean absolute error

ML machine learning

MPC model predictive control

MPM model-plant mismatch

MSE mean squared error

MV manipulated variable

NAN not a number

OLS ordinary least squares

OT operating technology

PCA principal component analysis

PDF probability density function

PID proportional-integral-derivative

P&ID process & instrumentation diagram

PLS partial least squares

PPV positive predictive value

PSE process systems engineering

PV process variable

RBF radial basis function

RNN recurrent neural network

RCC residual carbonate control

RELU rectified linear unit

RMSE root mean square error

SH squared-hinge

SIM subspace identification method

SMBO sequential model-based optimization

SME subject matter expert

SPM statistical process monitoring

SOTA state-of-the-art

SVC support vector classifier

SVM support vector machine

TEP Tennessee Eastman process

TN true negative

TPE tree-structured Parzen estimator

TP true positive

TPR true positive rate

UMAP uniform manifold approximation and projection

VARMAX vector auto-regressive moving-average exogenous input

Acknowledgments

Thank you to my supervisors Bhushan Gopaluni and Philip Loewen, as well as my committee members Yankai Cai, Paul Bicho, and Devin Marshman. I would also like to thank my industry collaborators Michel Ruel, Carole Prévost, Barry Hirtz, Travis Reinheimer, Cilius Van der Merwe, Peter Gorog, and Carl Sheehan, as well as my research colleagues, Sirish Shah, Ibrahim Yousef, Yiting Tsai, Qiugang Lu, Siang Lim, Liang Cao, and Nathan Lawrence. I would like to acknowledge BBA Engineering Consultants, Canfor Pulp, Spartan Controls, the National Science and Engineering Research Council of Canada, and the Izaak Walton Killam Memorial Fund for funding this research.

Finally, I would like to express my gratitude to my friends and family for supporting me throughout my graduate school journey. I would especially like to thank my parents and my wife Catherine. Without their support this work would not have been possible.

I think what is happening right now is not AI. That was an intellectual aspiration that is still alive today as an aspiration. I think this is akin to the development of chemical engineering from chemistry, or electrical engineering from electromagnetism. ... We have a proto field, which is statistics, more of the theoretical, algorithmic side of computer science. That was enough to start to build things, but what things? Systems that bring value to human beings, and use human data, and mix in human decisions. The engineering side of that is all ad hoc; that is what is emerging, in fact if you want to call machine learning a field, I think that is what it is, i.e., a proto-form of engineering based on statistical and computational ideas of previous generations.

— Michael I. Jordan (2020) [2]

Chapter 1

Introduction

More evidence-based decision-making in the operation, control, and optimization of large-scale industrial processes is critical for minimizing environmental impacts and maintaining economic viability. The adoption of statistics, computation, and machine learning (ML) in a world that is increasingly represented digitally, has led to more evidence-based decision-making across many walks of life [55]. *Process analytics* refers to the application of advanced analytics and ML techniques to manufacturing data [122]. This dissertation is about process analytics and ML with operating data from industrial processes (e.g., pyrometallurgy and kraft pulping) for advanced process control (APC) applications such as process monitoring, fault detection and diagnosis (FDD), and inferential sensing.

1.1 Machine Learning and Process Analytics

Throughout the last decade an immense amount of attention has been drawn to the field of artificial intelligence (AI). In addition to algorithmic and computational advances, a significant portion of that attention stems from a series of breakthroughs in a domain of AI known as ML. In fact, it is a sub-domain of ML, referred to as deep learning, that is responsible for a great deal of the recent interest around AI.

Deep learning is an emerging paradigm in ML that broadly involves using large

amounts of data to train artificial neural networks (ANNs) with multiple hidden layers, i.e., deep neural networks (DNNs). The diversity of parameter-expression architectures recently developed with deep ANNs is a modern technological achievement. Deep learning techniques have demonstrated state of the art results on pattern recognition problems [71]. While deep learning is commonly associated with pedagogical examples such as classifying handwritten digits (c.f., MNIST) or images of cats and dogs, it is in fact consequential to all three pillars of ML. As Figure 1.1 illustrates, the field of ML has three major pillars (i.e., paradigms), including supervised learning, unsupervised learning, and reinforcement learning [75, 124].

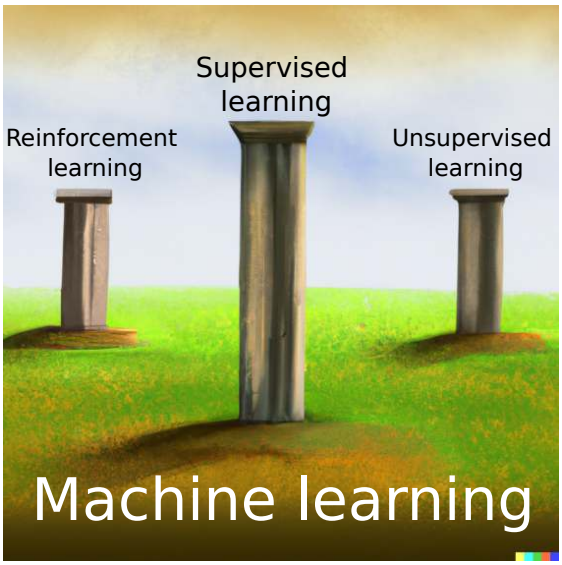


Figure 1.1: The field of machine learning includes three pillars, i.e., supervised learning, unsupervised learning, and reinforcement learning.

Reinforcement learning involves learning how an agent should interact with an environment to maximize a cumulative reward signal. This paradigm of ML has strong connections to optimal control as both involve optimizing a control policy by analyzing the feedback of interactions with a process or environment. Unsupervised learning involves learning unknown patterns and discovering hidden structures with unlabeled data. Common examples of unsupervised learning include data exploration activities such as clustering and certain forms of dimensionality

reduction, e.g., principal component analysis (PCA). Finally, supervised learning involves learning a mapping from input data to labeled output data.

Although unsupervised learning methods are studied for data exploration and dimensionality reduction, this work primarily involves using advanced techniques from supervised learning to increase the production efficiency and reduce the environmental impact of large-scale industrial processes. In supervised learning, matrices of feature vectors (or inputs) are presumed to provide contextual information for vectors of outputs (or targets). The term is inspired by viewing the labeled outputs as providing supervision to the ML algorithm [38]. For instance, given tuples of inputs $\boldsymbol{X}$ and outputs $\boldsymbol{Y}$, supervised learning algorithms aim to learn a mapping from $\boldsymbol{X} \rightarrow \boldsymbol{Y}$ with large quantities of labeled training data that accurately predicts $\boldsymbol{Y}$ given new observations of $\boldsymbol{X}$. Outputs are continuous for regression problems and discrete (or categorical) for classification problems.

In the context of kraft pulping, an example of regression is using historical process data to predict the specific energy of a lime kiln (i.e., the amount of energy that is expended per unit of lime produced). Alternatively, a relevant example of supervised learning for classification is learning a mapping from the historical process data to labeled process faults in order to predict future faults. In this classification example the output is discrete because it either indicates that there will be a fault (e.g., $Y = 1$) or that there will not be a fault (e.g., $Y = 0$). In fact, this predictive classification problem is the focus of the pyrometallurgical application presented in Chapter 3.

1.2 Motivation

Inference is the process of reaching conclusions based on evidence and reasoning. The ubiquity of advanced digital technology in society (e.g., smartphones) has highlighted the importance of inferential methods that can leverage large quantities of data. Whether it be through ML, optimization, or statistics, it has become more important than ever for engineers and scientists to enhance their inferential prowess. In fact, concepts such as AI, autonomous systems, and big data analytics have drawn increased interest for applications with chemical processes

[33, 99, 132]. While algorithmic progress on benchmark case studies has developed rapidly, successful industrial implementation has proven much more challenging. This research explores this discrepancy while addressing novel industrial case studies with historical operating data from pyrometallurgical and kraft pulping processes.

The focus of this work is primarily on supervised learning techniques for preliminary development of novel process monitoring solutions. This scope is largely inherited by the constraints of developing novel data-driven solutions to help mitigate outstanding process faults in safety-critical real-world industrial environments. To capture the full benefits of advanced automation, closed loop APC solutions are ultimately desirable. However, prematurely giving an innovative algorithm the ability to directly manipulate variables in an industrial process is typically neither appropriate nor responsible. The high-stakes nature of these environments often enforces a conservative solution development life-cycle. Implementing changes that impact production requires buy-in from multiple parties including management, engineers, operators, and subject matter experts (SMEs).

There are alternative, more theoretical, approaches to developing data-driven APC solutions. For example, if a high fidelity simulator (i.e., a digital twin) is available then innovative closed loop control algorithms can be implemented and evaluated in a low-risk environment. Process monitoring benchmark challenges can be used to compare many different algorithms on the same dataset. This flavor of research is often characterized by complicated ML workflows that seek incremental performance improvements on well-understood problems (e.g., benchmarks) for which many alternative solutions exist. Unfortunately, this style of research is of minimal practical utility and is over-represented in literature. The divide between the simplicity of simulations and the complexity of real-world processes is insufficiently addressed to yield practical utility. No coherent value proposition exists for implementing costly changes to a reliable, proven solution based on incremental performance in a low-fidelity experiment.

In a world captivated by the achievements of AlphaGo [117], the prospect of a deep reinforcement learning (DRL) controller with minimal dependence on process

knowledge has become highly enticing. As Chapter 2 describes, this prospect led to a torrent of research in advanced deep learning methods for FDD. However, thus far, the real-world impact of the deep learning revolution in the field of industrial process control has been relatively insignificant. Perhaps the prospect of giving a black-box DNN algorithm, with limited (if any) interpretability, the ability to control an oil refinery is more well-received at conferences than in control rooms.

There is value and a need for both practical, application-oriented research as well as theoretical, visionary research. Unfortunately, with the excitement around AI and Industry 4.0, a sensible balance of research styles has been neglected. There are citation incentive structures inherent in academia that reward broad intellectual trend-setting and early adoption of state-of-the-art (SOTA) algorithms. There are also challenges that disincentivize research with tangible aspirations such as intellectual property concerns with industrial collaborations, acquiring and learning from industrial process data, accounting for safety-critical process conditions, and obtaining buy-in from operators and SMEs. This dissertation aims to help researchers and practitioners overcome these challenges to deliver outcomes with strong prospects for industrial implementation and real-world impact.

Although this situation is untenable, it is not beyond repair. Embracing the need for ML researchers and practitioners to immerse themselves in the relevant application-specific process knowledge is essential. Instead of conveniently bypassing operator and SME concerns to develop SOTA solutions with minimal intention for industrial implementation, there is a need for prioritizing a commitment to do the hard work to improve industry outcomes. This involves rigorous consideration of process conditions, equipment limitations, and first principles process knowledge. It requires extensive collaboration with industry professionals to develop solutions that empower operators and SMEs to increase production efficiency.

1.3 Objectives

This dissertation contributes to the balance of theoretical and practical research in data-driven approaches to monitor, control, and optimize industrial processes. At a fundamental level, this research addresses two specific case studies and develops

process monitoring tools that use operating data to help mitigate outstanding faults in an electric arc furnace and a rotary lime sludge kiln. At a more abstract level, this work identifies underlying, process agnostic challenges encountered while learning from industrial process data. Guidelines for these challenges are provided by reflecting on experience from case studies and industry.

1.3.1 Improving operating outcomes with industrial data

The overarching objective of this dissertation is to help researchers and practitioners leverage historical operating data to improve operating outcomes. This objective is accomplished by studying best practices, proposing an overall methodology, demonstrating and refining it on case studies, and reflecting on underlying challenges to help others avoid common pitfalls. The proposed approach involves embracing domain-specific process knowledge and progressing from offline insights, to process monitoring, and ultimately closed loop control.

The proposed research methodology amounts to an APC solution development framework that prioritizes using historical process data to improve industrial outcomes. It involves learning from historical process data with ML, process knowledge, and SME guidance to develop and validate process monitoring tools that mitigate highly detrimental process faults. This framework provides a low-stakes proof-of-concept within the real operating environment to help novel process monitoring tools evolve into closed loop APC solutions.

Finally, given the popularity and success of benchmark challenges for research in ML and FDD, a novel FDD benchmark challenge is developed based on the pyrometallurgy case study. Large quantities of raw industrial operating data are published along with a novel supervised learning FDD challenge to improve the practical utility of process analytics benchmark research.

1.3.2 Predicting arc loss in an electric arc furnace

The objective of the first industrial case study is to predict the unexpected loss of plasma arc in an 80 MW twin electrode direct current (DC) electric arc furnace (EAF). This furnace serves as a smelter and is the most energy-intensive unit

operation in a 60,000 tonne/year integrated mining and pyrometallurgy process [63]. Specifically, the objective is to develop an algorithm that can use routine operating data to provide operators with a reliable early warning of an impending arc loss event (e.g., five minutes prior) such that preventative measures can be taken to avoid a significant process upset. As Chapter 3 demonstrates, the economic and environmental losses incurred by unexpected loss of plasma arc are immense. The novel FDD benchmark mentioned before is referred to as *The Arc Loss Challenge* [149]. The goal of publishing this benchmark is not just to enable more industrially relevant FDD research generally, but also to crowd-source enhanced solutions to the arc loss problem specifically.

1.3.3 Lime kiln monitoring with infrared thermal cameras

The problem formulation for the second industrial case study is not nearly as straightforward and there is no known means by which the data can be accurately labeled. Therefore, a significantly more immersive and rigorous consideration of process knowledge is undertaken. The objective of this research is to study how thermal cameras can be used with routine operating data and ML to improve the production efficiency of a rotary lime sludge kiln.

As the largest energy consumer in a 152,000 tonne/year kraft pulp mill it is of paramount importance, both from economic and environmental perspectives, to operate the kiln efficiently. Ring formation is known as the most troublesome problem for lime kiln operation [126]. As such, detecting and diagnosing ring formation is a primary focus of this research, and as Chapter 4 describes, the feasibility of this endeavor is highly dependent on data quality. In addition to kiln monitoring with thermal cameras, another focus of this research is the development of an inferential sensor to predict residual calcium carbonate and optimize lime kiln control.

1.4 Thesis Outline and Contributions

The remainder of this dissertation is divided into five chapters. Chapter 2 introduces advanced applications of ML for FDD, presents limitations in literature, highlights opportunities for leveraging industrial data, and identifies common underly-

ing challenges while learning from process data. Chapter 3 and Chapter 4 present the pyrometallurgy and kraft pulping case studies, respectively. Relevant literature is reviewed throughout these three chapters. Chapter 5 reflects on case study and industry experience to provide guidelines and best practices for overcoming the challenges encountered while applying analytics and ML to industrial process data. Finally, concluding remarks are provided in Chapter 6.

Chapter 2 provides background into industrial applications of process analytics and ML. Both academic literature and practical experience are leveraged to present advances in FDD, describe how value can be generated from historical process data, and propose an APC research methodology that focuses on improving operating outcomes with industrial data. This methodology is presented in Chapter 2 along with insights such as challenges and practical considerations informed by both literature and practice. Introducing the ML workflow terminology and procedures in Chapter 2 allows subsequent chapters to focus on the process knowledge and engineering required for each case study.

Chapter 3 presents the pyrometallurgical case study on loss of furnace plasma arc. First, important background information is provided to introduce the industrial process and the process fault. An end-to-end study is then presented that takes raw historical operating data, formulates an ML problem, develops various inferential sensor models using both traditional and advanced supervised learning methods, and then compares them. Practical lessons (e.g., the importance of process knowledge) are learned during this case study to inform better research practices.

The lime kiln case study is presented in Chapter 4 and it begins with a comprehensive introduction to the important role of rotary lime sludge kilns in the kraft pulping industry. The use of infrared thermal cameras on lime kilns is described and a novel data visualization strategy is presented. By analyzing aberrations in shell temperatures a previously unknown phenomenon known as rotational aliasing is discovered. Various methods for monitoring ring formation are studied. Inferential sensing of residual calcium carbonate is studied with various supervised learning methods to help optimize kiln energy efficiency.

Chapter 5 presents research findings and best practices for learning from industrial

process data. Challenges raised in Chapter 2 are addressed by guidelines discussed in Chapter 5. Reflecting on experience from the pyrometallurgy and kraft pulping case studies provides pitfalls and guidelines to help future researchers and practitioners. To conclude this chapter, existing benchmarks for FDD are reviewed, shortcomings are identified, and a novel industrial FDD benchmark is introduced.

Finally, Chapter 6 presents conclusions and future work specific to each of the industrial case studies. This is followed by providing high-level conclusions on improving the practical utility of research on applications of process analytics and ML for monitoring, optimization, and control of chemical and biological processes.

Chapter 2

Industrial Applications of Process Analytics and Machine Learning

"Applications of machine learning and artificial intelligence in advanced manufacturing and control of chemical and biological processes." Text to generate images with DALL·E 2 [1].[1]

[1]Note that these are computer generated images used only for abstract conceptual visualization.

This chapter provides background into industrial applications of process analytics and ML. A comprehensive literature review of all the APC use-cases for ML is well beyond the intended scope. Instead, Section 2.1 focuses on reviewing literature for FDD methods and recent advances with DNN architectures. Section 2.2 provides more pragmatic considerations. High impact use-cases are presented chronologically with respect to their role in the proposed APC solution development methodology. This methodology is further described by discussing the important challenges encountered during the various stages of learning from industrial process data.

2.1 Fault Detection and Diagnosis in Chemical and Biological Processes

The literature on ML and statistical methods for FDD is in some ways rich and in other ways limited. There are often only a limited number of studies on FDD methods for specific industrial applications (e.g., rotary kiln monitoring) and their relevance is often complicated by differences in the underlying process, lack of transparency, and technical limitations (i.e., missing context and incomplete data analysis). More generally, the FDD literature is quite advanced, particularly with respect to the breadth of advanced solution methods that have been proposed on popular benchmarks like the Tennessee Eastman process (TEP) [22].

As a well-developed field that has drawn industrial interest for over fifty years the terminology related to work in FDD can be overwhelming. For instance, FDD is sometimes considered a sub-discipline of process monitoring, or more specifically, statistical process monitoring (SPM). In other cases, FDD activities are considered a central component of abnormal event management (AEM), which is itself a key component of supervisory control. Knowledge of the different hypernyms for FDD is valuable for conducting a literature search, but otherwise it is of little importance relative to understanding the procedures and methods. Detection and diagnosis of process faults is not the complete picture as additional activities such as fault identification, fault isolation, fault estimation, fault reconstruction, and process recovery have been included in the general FDD procedure [22, 53].

This section introduces data-driven methods for fault detection and it includes ma-

terial from the following journal publication:

- L. D. Rippon, I. Yousef, B. Hosseini, A. Bouchoucha, J. F. Beaulieu, C. Prévost, M. Ruel, S. Shah, and R. B. Gopaluni. Representation learning and predictive classification: Application with an electric arc furnace. *Computers & Chemical Engineering*, 150:107304, 2021.

2.1.1 Process history based methods

Traditional methods for FDD have been comprehensively reviewed and broadly separated into three categories, i.e., i) quantitative model-based methods, ii) qualitative model-based methods, and iii) process history based methods [133–135]. The objective of learning from historical operating data implies a focus on process history based methods which are divided into quantitative and qualitative subcategories. Classical examples of quantitative process history based methods include popular ML methods such as PCA, partial least squares (PLS), the Bayes classifier, and early implementations of neural networks [134]. The PCA and PLS methods were fundamental to the development of FDD as a field so they are described briefly in what follows.

Principal component analysis

The PCA statistical procedure was introduced in the early 20th century to decompose a multivariate dataset into a basis set of linearly uncorrelated orthogonal variables called principal components [96]. It was subsequently used in multivariable quality control and has since been further extended and applied in process systems engineering (PSE) where it is categorized as a quantitative process history based method for FDD [46, 89, 125]. The convention for FDD is to calculate the Hotelling T^2 statistic with the largest singular values and the Q statistic with the smallest singular values. The T^2 statistic defines normal process behavior and any observation vectors that fall outside of the T^2 region indicate that a fault has occurred. Alternatively, the Q statistic is used to define a threshold that indicates whether or not the characteristics of the measurement noise have changed significantly [111, 116].

Consider a pre-processed set of historian data that has been centered (i.e., col-

umn means subtracted), $X \in \mathbb{R}^{n \times d}$ where X includes the output label data as additional columns. The covariance matrix of X is denoted $S \in \mathbb{R}^{d \times d}$ and is given by $S = X^\top X/(\boldsymbol{n}-1)$. The typical eigendecomposition of S is given by $S = V\Lambda V^{-1}$ where the i-th column of $V \in \mathbb{R}^{d \times d}$ is the eigenvector v_i of S or alternatively, the loading vectors or principal directions of the data X. The diagonal matrix $\Lambda \in \mathbb{R}^{d \times d}$ contains eigenvalues λ_i that are sorted in order of decreasing magnitude. Given a symmetric matrix S with distinct eigenvalues λ_i, the eigenvector columns of V are orthogonal (i.e., $V^{-1} = V^\top$) and the eigendecomposition becomes $S = V\Lambda V^\top$. The principal components or principal component scores can be calculated by projecting the data onto the principal directions, i.e., $C = XV$, where the i-th column of C is the i-th principal component of X [118]. Alternatively, PCA can be conducted with singular value decomposition of the centered data matrix X where singular values (σ_i) are related to eigenvalues by $\lambda_i = \sigma_i^2/(n-1)$ [116].

Partial least squares

As with PCA, PLS (also known as projection to latent structures) is a linear representation learning method with a rich history of use in PSE. The PLS approach was first introduced by Herman Wold in the 1970s and has since been used extensively in chemical process industries as a chemometrics method for applications such as FDD [72, 141]. One drawback of PCA is that although some principal components may describe significant variance in X, those same principal components might not be relevant for predicting the output labels, Y. As a supervised learning method, PLS regression maximizes the covariance between the input data, X, and output data (or labels), $Y \in \mathbb{R}^{n \times d_y}$, in the latent space via the non-linear iterative PLS algorithm [113].

The centered input matrix X and output matrix Y are each decomposed as, $X = LP^\top + E$ and $Y = MQ^\top + F$, where $L \in \mathbb{R}^{n \times a}$ and $M \in \mathbb{R}^{n \times a}$ are latent score matrices, $P \in \mathbb{R}^{d_x \times a}$ and $Q \in \mathbb{R}^{d_y \times a}$ are loading matrices, $E \in \mathbb{R}^{n \times d_x}$ and $F \in \mathbb{R}^{n \times d_y}$ are residual matrices and a is the PLS component or reduction order [112]. The iterative PLS regression algorithm initializes $X_1 \coloneqq X$ and $Y_1 \coloneqq Y$ and then proceeds to maximizing $l_i^\top m_i$ (for each iteration i) by initializing m_1 as one column of Y and

solving the following set of equations until convergence is achieved:

$$w_1 = \frac{X_1^\top m_1}{\|X_1^\top m_1\|}, \quad l_1 = X_1 w_1, \quad q_1 = \frac{Y_1^\top l_1}{\|Y_1^\top l_1\|}, \quad \text{and} \quad m_1 = Y_1 q_1, \tag{2.1}$$

where $\|\cdots\|$ represents the Euclidean norm or ℓ_2 norm. The X-weights (w_1) are updated with the Y-scores (m_1) until the change in l_1 is negligible or below some specified error [112, 141]. The same procedure is repeated for the next iteration by replacing X and Y with the residual matrices, i.e., $X_{i+1} = E_i = X_i - l_i p_i^\top$ and $Y_{i+1} = F_i = Y_i - m_i q_i^\top$ where $p_i = X_i^\top l_i / \|l_i^\top l_i\|$.

Qualitative process history based methods

Contrary to quantitative process history based methods are qualitative process history based methods such as expert systems. An expert system is a computer program that emulates the reasoning ability of an SME with a knowledge base and if-then rules-based logic [7]. Combining expert systems with ML enables the integration of first principles process knowledge with prediction models. As an interesting aside, expert systems are an example of symbolic AI which was a dominant paradigm in AI research before the contemporary preeminence of deep learning [38] and other connectionist approaches to AI.

2.1.2 Performance benchmarking - limitations and opportunities

Restricting scope to process history based methods is helpful, but more insight can be gained by applying these methods on standard benchmark datasets and comparing their performance. The TEP benchmark dataset is particularly relevant as it has a well established history in process control and FDD research. As helpful as benchmark challenges can be (e.g., for literature review, hands-on practice, and model evaluation) they also have limitations and can be prone to misuse.

Chapter 1 included some critical remarks on the abundance of research that uses benchmark datasets to circumvent the more tedious and less academically exalted aspects of applying ML solutions to industrial problems. The procedure for addressing an industrial problem with an ML or data-driven solution broadly follows

the stages outlined in Figure 2.1[2]. Research with simulated (or highly procured)

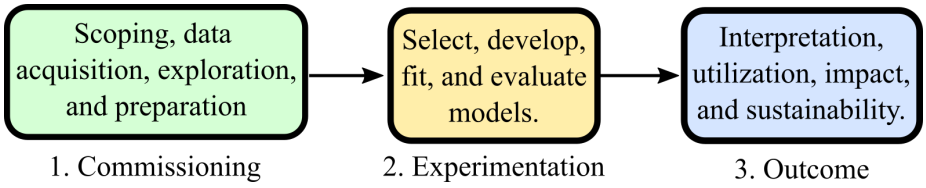


Figure 2.1: Stages involved with delivering a novel data-driven solution.

benchmarks can focus almost solely on the second stage (i.e., experimentation) which involves developing a novel algorithm, fitting training data, evaluating, and updating parameters iteratively until desired performance is achieved. Conversely, efforts towards industrial solutions require significant emphasis on the first stage of onerous, but necessary preparation, and the difficult third stage of interpreting results and delivering outcomes.

One drawback of a focus on benchmark oriented research is a misalignment between the objectives of the benchmark and the requirements for industrial impact. Practical operation requires solutions that provide long-term robustness and safety while benchmark studies often prioritize performance in a relatively limited environment. Implementing robust solutions requires domain knowledge and an understanding of the circumstances that lead to degradation of model performance such as equipment fouling, measurement disturbances, and confounding variables. To complement the existing literature, this research emphasizes the first and third stages of Figure 2.1 while developing novel solutions for case study applications.

Another pitfall of benchmark datasets is the production of research with minimal scope beyond the benchmark challenge. This can result in literature that gives the false impression of innovation and technological progress. Research with clean simulated data is dominant in literature. One explanation is that there exists a strong bias towards the experimentation stage in Figure 2.1 with respect to publishing standards [138]. In the context of complex industrial processes, pragmatic reasons such as the difficulty to obtain and procure historical operating data may offer further explanations. However, real-world impact and industry adoption should

[2]This low resolution procedure will be expanded in Section 2.2.

be a primary focus rather than benchmarking for the sake of benchmarking, i.e., publishing incremental performance improvements on simulated data with minimal intent or consideration for utilization.

When used correctly, benchmark datasets can play a valuable role in providing a consistent evaluation environment that enables systematic comparison of different modeling methods. There is a significant opportunity in FDD research for a modern benchmark challenge with raw operating data from a large-scale industrial process and accurate labels of a process fault. In fact, the introduction of such a benchmark, specifically *The Arc Loss Challenge*, is one of the contributions associated with the case study presented in Chapter 3 [149]. However, the consistency of benchmark evaluations is leveraged here by focusing on literature associated with FDD methods that are applied to the TEP benchmark.

The popular TEP is often described in literature as an industrial FDD benchmark, but more accurately it is a simulated benchmark that is modeled after a real industrial process. Although there are different versions, the TEP commonly involves five simulated unit operations including a reactor, a stripper, a condenser, a compressor, and a separator [9, 29]. The benchmark problem consists of detecting and diagnosing 21 different faults caused by events such as step changes, random disturbances, slow drift, actuator stiction, and unknown factors. Researchers can evaluate their FDD techniques on a dataset containing 52 variables, including 41 process variables (PVs) and 11 manipulated variables (MVs). Standard training and testing sets have been made available that include data from each fault along with normal operating data to enable systematic evaluation of binary (i.e., detection) and multiclass (i.e., diagnosis) classification algorithms. The TEP is also used as a benchmark for control, system identification, and alarm management.

2.1.3 Deep learning for fault detection

In the past two decades research into quantitative process history based methods for FDD has developed rapidly and an exhaustive account of this progress is beyond this scope. Roughly one decade ago a highly cited study was published that compares traditional FDD methods such as PCA, PLS, independent component anal-

ysis (ICA), modifications of these, and a subspace identification method (SIM) on the TEP benchmark. This study highlights the higher fault detection rates (FDRs), lower computation cost, and minimal assumptions regarding the process data as positive features of the SIM [147]. However, throughout the last decade an immense amount of attention has been drawn to AI from a series of breakthroughs in a sub-discipline of ML known as deep learning.

Deep learning is an emerging paradigm in ML that involves using large amounts of data to train ANNs with multiple hidden layers and diverse architectures in order to obtain more expressive function approximations. Recently, deep learning techniques have dominated research in many scientific disciplines including the development of quantitative process history based methods for FDD. One of the first investigations into deep learning for FDD resulted in a dramatic improvement on the average FDR for the TEP benchmark relative to the SIM [81]. Before introducing more advanced deep learning architectures, some background into ANNs is provided in what follows.

Artificial neural networks

Conceptually, ANNs were inspired by the structure and function of neurons in the human brain [50]. Neural networks have undergone at least three historical waves of popularity beginning with cybernetics in the mid 20th century, connectionism in the late 1900s and the current manifestation of deep learning that began in 2006 [38]. The deep learning wave of popularity resulted from a breakthrough in the efficiency of training deep networks by Geoff Hinton's research group, referred to as greedy layerwise unsupervised training [10]. The versatility and non-linear representation capacity of ANNs has drawn immense interest from the scientific community as a classifier for modeling complex relationships [130].

The perceptron, introduced by the psychologist Frank Rosenblatt, is the first and most simple example of a modern neural network that was explicitly used for binary classification of linearly separable functions [110][57]. The perceptron is a building block for complex multi-layered ANNs that involve input layers, hidden layers, and output layers consisting of neurons connected with learned weights

[24]. Linear combinations of inputs and weights are fed to nonlinear activation functions at multiple layers to provide a versatile, nonlinear model [93].

Activation function selection can have a significant influence on ANN performance [28]. Types of activation functions include sigmoid, rectified linear unit (RELU), softmax, and many more. The non-linear output from the activation function is represented by:

$$Z = (\sum_{i=1}^{n} x_i w_i + b) \qquad y = f(Z), \tag{2.2}$$

where $x_1, x_2, \ldots, x_n$ represent the n inputs of the perceptron, $w_1, w_2, \ldots, w_n$ are the weights given to the respective input, b is a bias term, and f represents the chosen activation function for this layer. Back propagation is used to calculate the gradient of the output error with respect to the neuron weights and gradient descent is used to train the network to minimize the output error [69].

Deep learning architectures

Deep learning tackles the problem of representation learning by using complex neural architectures to generate nested representations that are functions of simpler representations [38]. Predominant deep learning architectures such as generative adversarial networks (GANs), autoencoders (AEs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs) have all been adopted to develop advanced strategies for FDD. Figure 2.2 illustrates some unique features of these neural architectures while providing a summary of literature involving their application to the TEP benchmark. Citations in Figure 2.2 demonstrate the abundance and variety of proposed deep learning FDD solutions on the TEP.

The GAN architecture, inspired by game theory, consists of a generator network (G) that is trying to deceive a discriminator network (D). The AE architecture, often used for dimensionality reduction, consists of an encoder network (E) that learns a latent representation of the inputs and a decoder network (D) that attempts to reconstruct the inputs by sampling the latent representation. The CNN architecture learns a set of convolutional filters that extract features from the input data in a hierarchical fashion. Finally, the RNN architectures, including variants such as gated recurrent unit (GRU) and long short-term memory (LSTM) networks, have one

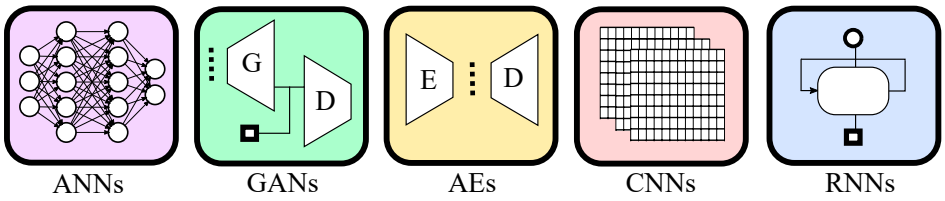


Figure 2.2: A summary of deep learning techniques for FDD on TEP benchmark dataset including feedforward ANNs [43], GANs [34, 52, 74, 120, 145], AEs [21, 51, 81, 82, 144, 150, 154], CNNs [18, 20, 35, 101, 142], and RNNs [19, 123, 152, 153, 155].

or more internal states that function as a memory to encode temporal information.

Many of the deep learning FDD methods that are cited in Figure 2.2 use advanced variants (e.g., bidirectional RNNs [19]) and combined architectures (e.g., recurrent AEs [21]). Moreover, some of these methods combine deep learning with other ML methods such as *k*-nearest neighborss (*k*NNs) and random forests. Standard feedforward ANNs are also represented in Figure 2.2 and they have been shown to provide impressive FDD results on the TEP benchmark [43].

Multiple deep learning architectures are applied in the case studies presented in Chapter 3 and Chapter 4. The focus of this work is not a theoretical review of deep learning, but given the relevance of the CNN architecture for both case studies, a brief digression is provided in what follows.

Convolutional neural networks

State of the art performance has been achieved using CNNs on object recognition and natural language processing tasks [10]. In the late 1980s CNNs were introduced to address visual pattern recognition problems such as handwritten digit recognition [66]. Instead of exclusively using fully-connected layers, CNNs use local connections (i.e., local receptive fields) to extract elementary features which are then combined by subsequent layers in a hierarchical feature extraction procedure [67]. Convolution with a kernel whose weights are learned through back-propagation creates the local receptive field for each feature map [70]. The receptive field and the dimensions of the resulting feature map are governed by the size of the kernel

and the stride that the kernel takes over the input image (or input feature map). For a single input image, the number of output feature maps after the first convolution layer is equivalent to the number of learnable kernels specified for that layer.

For convolutional layer l, the output of the jth feature map is given by [16]:

$$\boldsymbol{x}_j^l = f\Big(\sum_{i \in M_j} \boldsymbol{x}_i^{l-1} * \boldsymbol{k}_{i,j}^l + b_j^l\Big) \tag{2.3}$$

where M_j is the set of input feature maps, $*$ is the discrete convolution operation, $k_{i,j}$ is the learnable kernel from input map i to output map j, and b_j is the additive bias for output map j.

It is common to follow convolution layers with a sub-sampling procedure known as a pooling layer. The output feature map is sub-sampled to create a lower dimensional feature map by applying a receptive field that converts the output at a certain location to a summary statistic of nearby outputs [38]. Two common types of sub-sampling operations are max pooling and average pooling. Pooling layers can help to improve computation and prevent overfitting [142]. As Figure 2.3 illustrates, the output of the last pooling layer is flattened before being passed to a fully-connected network.

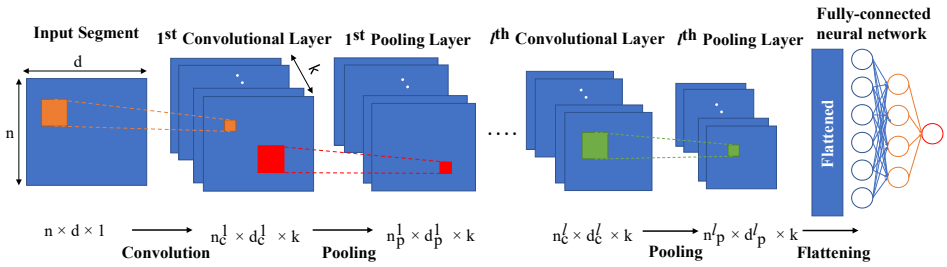


Figure 2.3: Illustration of the CNN predictive classifier architecture [108].

Deep learning is a rapidly evolving field so it is expected that many FDD algorithms will continue to be proposed based on the SOTA of deep learning and tested on the TEP benchmark. Unfortunately, there are also shortcomings in the TEP literature that prevent a concise empirical takeaway. Notable shortcomings include the use of

different datasets, evaluating subsets of faults, using different evaluation metrics, and most importantly a lack of commitment to transparency and reproducibility (with some exceptions [155]). Nevertheless, the literature that contributes variations of deep learning methods applied to the TEP benchmark is plentiful.

This dissertation aims to contribute by studying and developing novel applications for specific operating problems to drive better industry outcomes. Although the selection of powerful classification and regression models is important, the use of domain knowledge to identify ML opportunities and develop novel (and practical) solutions is significantly more substantive. The remainder of this chapter discusses more pragmatic aspects of learning from industrial process data to improve operating outcomes.

2.2 Learning from Industrial Process Data

The surface of ways to generate value from industrial process data has hardly been scratched. With the onset of a societal transformation driven by AI and ML, this subject is ripe for creative minds to find innovative ways to improve manufacturing outcomes by learning from this highly underutilized historical data. This section provides practical considerations for learning from industrial process data. Section 2.2.1 presents different types of use-cases and ties them together into a recommended strategy for APC solution development. Section 2.2.2 discusses challenges encountered by practitioners while also introducing the overall ML workflow for industrial process analytics.

2.2.1 Improving operating outcomes with process analytics

There are many different ways that operating outcomes can be improved by learning from industrial process data. Specifying the data use-case can help determine the data preparation and process analytic methodologies. This subsection describes three broad use-cases for learning from process data including offline investigations, process monitoring, and closed loop control. A creative mind can conceive of many more use-cases for this data, but these stages are sufficient to introduce the proposed APC solution development strategy. Developing data-driven APC so-

lutions with a high potential for impact involves deriving valuable insights by investigating offline data, improving operator capabilities with process monitoring, and ultimately directly controlling the process with closed loop algorithms.

Offline investigations

Offline investigations of historical process data provide many opportunities to generate valuable operating insights. Research activities such as feasibility studies and data-driven opportunity estimation can improve evidence-based reasoning for resource allocation. Productive allocation of capital to upgrade equipment, instrumentation, control strategies, and/or operator training is critical for long term production efficiency and sustainability.

Offline analysis of historical process data can enable accurate environmental assessments to ensure regulatory compliance. Historical process data can be studied to provide insights such as key performance indicator (KPI) benchmarking with industry peers. Troubleshooting investigations into process upsets rely on historical data for root cause analysis. Operating insights such as production losses and reliability issues can be identified through offline investigations of process data.

A simple demonstration is provided here to show how industrial process data can be used to characterize production losses. Consider a continuous process running at full production (i.e., highest sustainable capacity). Economically it is typically ideal to operate this capital intensive process at full production all of the time (i.e., 24 hours per day, 365 days per year). In reality, this is not possible, as extended plant shutdowns are necessary for maintenance. Moreover, unplanned downtime and short-term upsets are responsible for considerable production losses. Insights about the proportion of lost production attributable to each type of upset can help identify opportunities. This demonstration analyzes production-dependent variables to quantify losses based on upset duration. Specifically, for the pyrometallurgy process the furnace feed, furnace power, and calciner feed are considered whereas for the kraft pulping process the mud filter feed rates, kiln feed, and kiln fuel flow are analyzed.

An algorithm is developed that uses self-referencing definitions of *a process up-*

set and *full production* in order to quantify the production losses attributed to the following three categories of upsets: i) long duration upsets (i.e., greater than five days), ii) medium duration upsets (i.e., less than five days and greater than one hour), and iii) short duration upsets (i.e., less than one hour). These simple duration thresholds provide valuable diagnostics for overall process health because they are indicative of different underlying symptoms. Long duration upsets capture lost productivity from scheduled shutdowns whereas medium duration and short duration upsets capture different types of process faults. Short duration upsets can also include spurious production losses from instrumentation issues (e.g., aliasing).

The percentage of lost production across the different upset duration categories is presented for the pyrometallurgy variables in Figure 2.4 and for the kraft pulping variables in Figure 2.5. Five minute time-averaged samples are used for both datasets. The kraft pulping data includes over five years worth of data whereas the pyrometallurgy variables include only one year of operating data.[3]

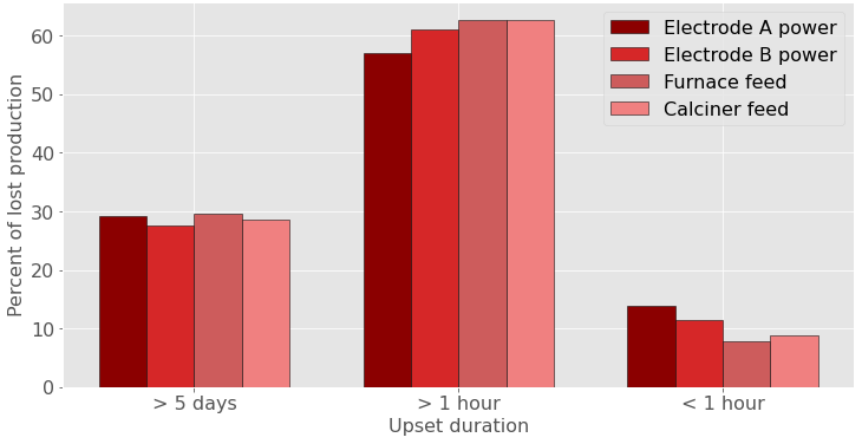


Figure 2.4: Lost production due to upsets of varying duration for the furnace feed, furnace power, and calciner feed in a pyrometallurgical process.

Insights can be obtained by using domain knowledge to compare differences in the distribution of losses both across unit operations within a process, and across

[3]This is important to consider while interpreting results because one year may be insufficient for an accurate representation of long duration process upsets.

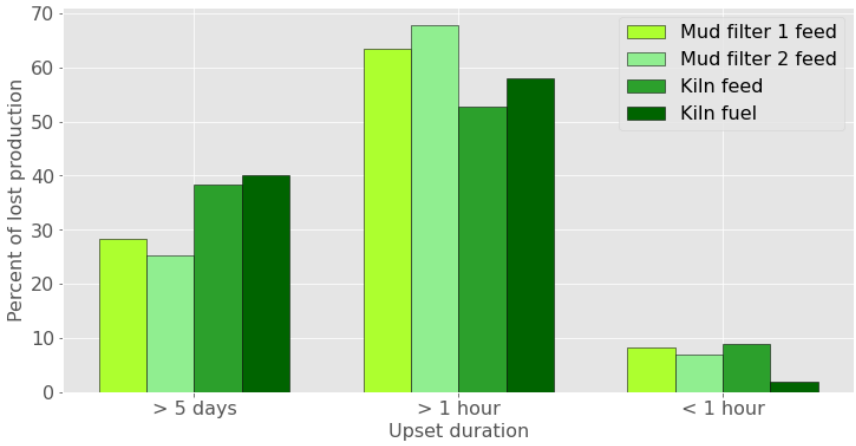


Figure 2.5: Lost production due to upsets of varying duration for the mud filter feeds, kiln feed, and kiln fuel usage in the recausticizing area of a kraft pulp mill.

multiple processes. For example, the proportion of short duration losses is lower for the kiln fuel relative to the feed rates because the mud precoat filters must be cleaned periodically with sheet drops that temporarily upset the feed rates. Continuous precoat renewal (CPR) systems have been developed specifically to minimize these sheet drop upsets and their associated production losses. Another example of a symptom is the relatively high proportion of short duration furnace power losses in Figure 2.4. This is indicative of the arc loss fault occurring in the furnace as described in Chapter 3.

This type of data-driven analysis can be made more rigorous, expanded in scope, and augmented by further investigations (e.g., data visualization). More importantly, the improvement iteration cycle can be shortened by increasing the rate of feedback at which these insights can influence operation. Online (i.e., real-time) implementations can provide more direct operating benefits by enhancing operator capabilities with process monitoring and fault detection tools.

Process monitoring and fault detection

Process monitoring and fault detection algorithms are differentiated from offline investigations by analyzing online data to provide time-sensitive insights that improve operation. Many of the data analytics and visualization strategies that are useful for offline investigation can be adapted for online implementation. Enhancing data visualization with domain-specific analytics is a valuable and underappreciated technique for developing process monitoring tools. More generally, equipping SMEs with the ability to interpret large quantities of process data in a user-friendly and intuitive manner, is an often overlooked means of generating value from process data. For example, during the investigation of the lime kiln case study (presented in Chapter 4) a novel shell temperature visualization strategy is developed for offline investigation of ring formation [103]. This visualization technique has now been successfully adapted into a commercial product for monitoring ring formation in lime kilns with real-time shell temperature data [14].

In addition to helping operators visualize process conditions, process monitoring and fault detection applications can include sophisticated ML algorithms for use-cases such as soft sensor development, model-plant mismatch (MPM) detection, and fault prediction. For example, a fault prediction algorithm is developed (as described in Chapter 3) to mitigate the recurring arc loss fault. By providing operators with a warning of an impending arc loss event the operators can take corrective actions and minimize production losses.

Process monitoring and fault detection applications are an ideal subject for academic research. They can be developed, investigated, and validated with offline process data and they can be implemented in live operating environments with minimal risk of upsetting operations. Although these applications generate insights from online data, they require operators to close the feedback loop and take actions that improve the process. Closed loop process control methods are differentiated from process monitoring methods by using measurement feedback to directly control the process. Inferential sensor applications (e.g., residual carbonate prediction) can be deployed initially as process monitoring methods before graduating to closed loop control strategies once their predictions are proven to be reliable.

Closed loop process control

Closed loop process control applications involve direct interaction by sensing and manipulating the state of the process to meet control objectives with minimal operator interference. Closed loop control applications can have varying levels of complexity such as simple regulatory control loops, supervisory control, and APC applications.

The ability to directly manipulate the state of the process gives closed loop control algorithms an immense potential to create value from industrial process data. Conversely, faulty applications of closed loop control can cause lost production, environmental damage, and safety incidents. This risk-reward distribution is why closed loop APC solutions are both highly coveted and difficult to innovate.

The pace of innovation in practice is far more modest than the pace observed in academic research which is often limited to (and developed for) simulated environments. Although many sophisticated algorithms (e.g., DRL) have been proposed, over 90% of industrial control systems still use proportional-integral-derivative (PID) control algorithms that were introduced in the early 1920s [11, 91].

Data processing requirements for closed loop control applications focus on safety, security, and reliability within the constraints of the distributed control system (DCS). Reliability in this context implies both reliable connectivity with real-time data as well as reliable predictions (e.g., setpoint changes) from robust models. Legacy operating technology (OT) systems are common among Canada's aging manufacturing facilities which can create data connectivity challenges for advanced closed loop control solutions.

The research presented in this thesis primarily contributes to the first two use-cases, i.e., offline investigation and process monitoring. However, the inferential sensing applications introduced in the following chapters can develop into closed loop APC solutions with further validation in production. Increased synergy between industry and academia is essential for better innovation in closed loop control research with practical significance. The remainder of this chapter explores considerations and challenges associated with learning from industrial process data.

2.2.2 Challenges while learning from industrial process data

Many permutations of options are available for constructing the data workflow of tasks such as data acquisition, preparation, modeling, and evaluation. Practical utility, sustainability, maintenance, and cost are important considerations to weigh against workflow complexity. Rather than attempt to prescribe a generalized optimal data workflow, this section presents practical considerations and challenges for process industry applications. In part because these tasks are critical and underappreciated, but also to provide background to enable subsequent chapters to focus on applications. The challenges introduced here are revisited in Chapter 5 after presenting the two industrial case studies.

Identifying and framing a data-driven opportunity

Identifying and framing a data-driven opportunity is fundamental to the conception and success of process analytics and ML applications with industrial process data. The following steps and associated challenges are important to consider during the initial conception of industrial process analytics applications:

- **Impact and benefit estimation:** is important to motivate stakeholders and justify expenditure of time and resources. Is there a reasonable expectation for the proposed research application to provide significant benefits? What types of benefits are valued by stakeholders?
- **Alignment on objectives and resource requirements:** with research partners and industrial collaborators. Research with industrial data can be limited by insufficient stakeholder alignment on the objectives of the study and a realistic view of the resources required to achieve them.
- **Availability and integrity of a source of ground truth:** to provide reliable validation and supervisory feedback for ML. The success of supervised learning algorithms is highly dependent on the integrity of the output labels.

These are just a subset of the many important factors to consider in order to avoid undertaking an ill-posed data-driven APC research project.

Data acquisition and contextualization

Data acquisition involves logistical challenges that have been historically difficult to overcome due to insufficient digital plumbing. This situation is rapidly improving. Comprehensive data contextualization is an often neglected task that is critical for analysis and opportunity identification.

- **Establishing connections with data sources:** to ensure efficient collaborations and to satisfy the evolving requirements of developing a novel solution.
- **Selecting the industrial data:** includes determining which variables to acquire and how to retrieve them, e.g., periods of operation, sampling frequencies, interpolation, and data compression.
- **Exploratory analysis and data visualization:** involves challenges such as handling overwhelming amounts of data, understanding the complexities of the underlying process, and using (or developing) tools for data visualization.
- **Data contextualization:** is critical for preparing data, developing models, and interpreting experimental results. Acquiring supplemental information to contextualize data involves consulting industry collaborators and SMEs.
- **Shortcomings of industrial data:** are an unfortunate reality that must be acknowledged and addressed while learning from industrial process data.

Thoughtful acquisition and contextualization of industrial process data provides an important foundation for downstream data-driven research activities.

Data preparation

Preparing the data for statistical learning is one of the most difficult and time consuming aspects of learning from industrial process data.

- **Data validation and reconciliation:** is necessary to ensure the quality of the data is sufficient for the desired objectives.
- **Data structuring:** involves challenges associated with aligning data with different measurement configurations. These challenges include labeling (or

selecting) target variables and sampling corresponding features.

- **Data cleaning:** of raw industrial operating data poses many challenges such as filtering irrelevant periods of operation and efficiently addressing invalid data.
- **Feature engineering and representation learning:** can impact the performance and computational feasibility of proposed solutions. Challenges include addressing redundant variables, standardizing features, engineering features with domain knowledge, and learning data-driven representations.

Process knowledge is important for many data preparation tasks such as data validation, segmenting data, output labeling, cleaning the data, engineering features, and designing visualization strategies.

Method selection, development, and evaluation

The selection, development, and evaluation of modeling methods receives significant attention in literature. However, practical challenges and limitations that are critical for industrial adoption, are often dismissed.

- **Storing and loading industrial process data:** efficiently is important for model development and it can be challenging due to the size and sensitive nature of the raw data.
- **Selecting suitable methods:** can involve challenges such as considering the available data and the intended application, as well as balancing competing interests such as academic novelty and industrial constraints.
- **Metrics that align with outcomes:** chasing irrelevant model metrics at the expense of business and environmental outcomes is a common pitfall encountered during model development.
- **Conducting comprehensive experiments:** involves challenges with organization, documentation, and consistent implementation to ensure the experimental results are reproducible.

Unfortunately, method selection in process analytics research has diverged significantly from method selection in industrial environments. A unified focus on improving operating outcomes can help reconcile this gap.

Deploying sustainable solutions

Soft sensors and other novel APC applications can struggle to provide sustained value in industrial production environments after initial commissioning. The following list includes important considerations for deploying sustainable solutions:

- **Understanding trends in utilization:** and identifying factors that lead to poor utilization is a challenging task that is necessary for obtaining sustained benefits.
- **Robustness, durability, and model maintenance:** are key considerations for implementing sustainable data-driven solutions in industrial environments.
- **Neglecting end-user experience:** is a common pitfall that can inhibit user adoption and prevent solution deployment.

In the following chapters two distinct industrial case studies are presented. The challenges listed here are encountered during these case studies. The lessons learned from these case studies are reflected upon to inform the discussion of guidelines and best practices presented in Chapter 5.

Chapter 3

Predicting Arc Loss in an Electric Arc Furnace

"A plasma arc provides heat for smelting in an industrial electric arc furnace at a pyrometallurgical plant." Text to generate images with DALL·E 2 [1].[1]

[1] Note that these are computer generated images used only for abstract conceptual visualization.

This case study addresses a novel industrial fault detection problem from both traditional and contemporary approaches to process analytics. Traditional approaches such as PLS are compared with techniques inspired by deep representation learning such as CNNs. Contributions include the formulation and introduction of a novel industrial predictive classification problem, the design and implementation of a comprehensive ML workflow that converts raw industrial data into critical operational insights, and the presentation of a robust comparative analysis between traditional and contemporary approaches to representation learning and binary classification.

Specifically, this work addresses the unexpected loss of plasma arc in the EAF that serves as a smelter in a large-scale metallurgical process refine ores into base metals. The objective is to learn an efficient and informative representation from the raw industrial data that enables the prediction of an arc loss event such that operators can take corrective actions. A comprehensive representation learning and predictive classification framework is presented for development of the inferential sensor from large quantities of historical industrial process data.

The majority of this chapter covers the formulation of the arc loss process fault as a supervised learning problem, preparing the industrial process data, and developing the soft sensor. This material is based on the following publications:

- L. D. Rippon, I. Yousef, R. B. Gopaluni, B. Hosseini, J. F. Beaulieu, C. Prévost, and S. L. Shah. Process analytics and machine learning to predict arc loss in an electric arc furnace. In *59th Conference of Metallurgists 2020 hosting the 4th International Uranium Conference*, 2020.

- L. D. Rippon, I. Yousef, B. Hosseini, A. Bouchoucha, J. F. Beaulieu, C. Prévost, M. Ruel, S. Shah, and R. B. Gopaluni. Representation learning and predictive classification: Application with an electric arc furnace. *Computers & Chemical Engineering*, 150:107304, 2021

The final section introduces recent work to publish the data and problem formulation as a modern benchmark challenge for fault detection with ML on industrial data. This work has been accepted for publication as follows:

- I. Yousef, L. D. Rippon, C. Prévost, S. L. Shah, and R. B. Gopaluni. The

arc loss challenge: A novel industrial benchmark for process analytics and machine learning. *Journal of Process Control*, 128:103023, 2023

3.1 Introduction to Arc Loss in an Electric Furnace

Consider an industrial mining and metallurgy operation that continues production until a fixed deposit of saprolite ore has been extracted and processed. In this context, assuming a carbon intensive energy source, it is imperative to minimize the expended energy per unit of on-spec product in order to minimize the environmental footprint over the lifetime of the operation. This work helps to support this objective by investigating the use of techniques from both traditional process analytics and advances in ML to predict the onset of arc loss such that operators can take remedial actions and maintain efficient operation.

Stable smelter operation is critical for successful production of base metals from particulate ore. This work studies the operation of an industrial DC EAF that operates as a smelter in a large-scale pyrometallurgical process. Specifically, unexpected loss of the plasma arc is an important unresolved problem with a significant impact on the production efficiency of the process. Moreover, given that EAFs are highly energy intensive units, even minimal improvements to the overall production efficiency represent meaningful reductions in the environmental footprint of the process over the lifetime of operation. To reduce the overall environmental footprint a predictive inferential sensor is proposed to identify high risk operating regimes. Once a high-risk situation is identified the alarm instructs operators to take corrective actions to avoid the loss of arc.

A simplified depiction of the broader mining and metallurgical operation is illustrated in Figure 3.1. Drilling is performed to characterize the deposit and ore is extracted from an open pit mine using trucks and hydraulic shovels. Dump trucks transport the ore to the ore preparation plant where it is screened to remove waste rock and crushed to prepare it for further processing. The crushed ore is then conveyed to the metallurgical plant (green) which is the process of interest for this work. Initial operations in the metallurgical plant are milling and drying which are conducted using hammer mill flash dryers. Dried ore is fed to a series of cal-

cining cyclones where it is dehydrated before being sent to fluidized bed reducers to remove oxides and improve the electrical efficiency of the subsequent smelting operation in the DC EAF [54, 60, 85]. The DC EAF is the operation that exhibits the unexplained loss of plasma arc fault.

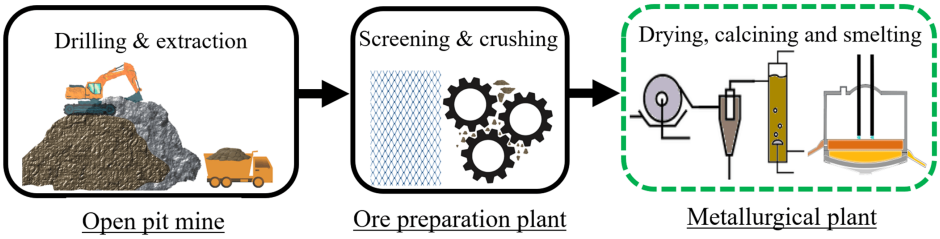


Figure 3.1: Simplified illustration of the relevant mining and metallurgical processes [106].

Through the upstream processing stages the raw ore is converted to a fine particulate feed that is fed to the DC EAF for smelting. As depicted in Figure 3.2, an open plasma arc spans from the cathode to the anode providing energy required to maintain temperatures of the slag and alloy above $1400^{\circ}C$, depending on the composition of the slag [65]. The cathode consists of two hollow graphite electrodes while the anode is the molten slag. The roof and side walls of the furnace are water cooled whereas the bottom anode is air cooled to maintain safe structural temperatures [47, 48]. Multiple ports along the roof feed the furnace while the slag and alloy are tapped from the launders intermittently [62]. Hot off-gas released from the furnace is recycled to provide upstream preheating. This work is directly relevant to a variety of EAF operations including nine in the Canadian steel-making industry [92].

As an energy intensive unit, it is critical to operate the EAF in a stable manner to maximize production efficiency. Unexpected loss of the plasma arc is a recurring and unresolved fault that significantly impacts the production rate and the electrical efficiency of the furnace. There are three primary categories of suspected arc loss mechanisms, i.e., electrical disturbances from the DC power supply, feed disturbances from the upstream metallurgical processes, and the operation of the EAF. Therefore, a broad process aspect ratio is considered in the representation learning

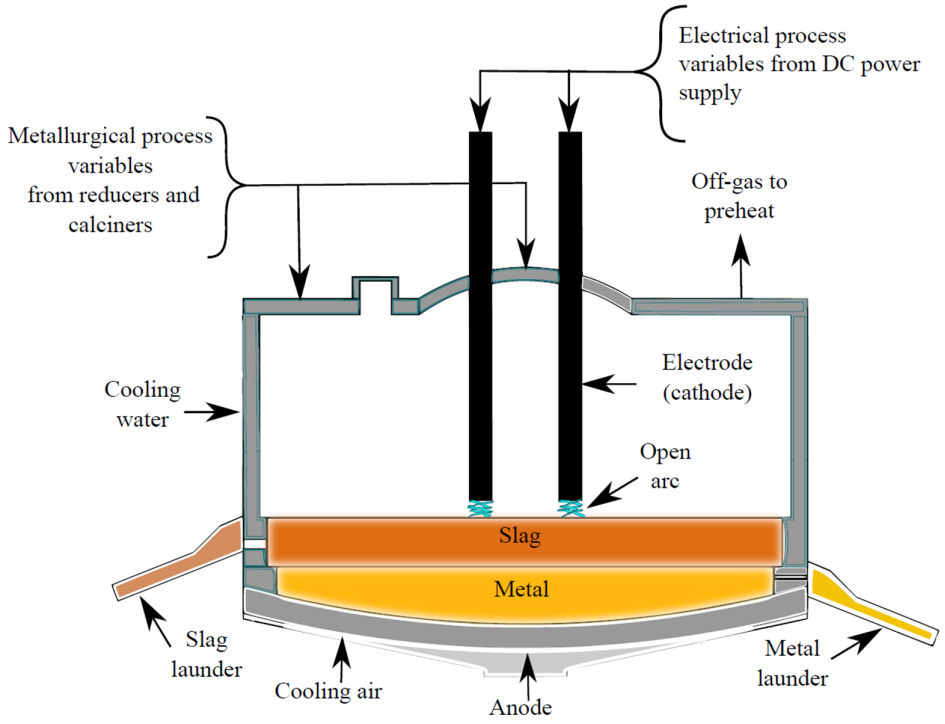


Figure 3.2: An illustration of a direct current electric arc furnace [106].

analysis that includes dozens of MVs from the power supply, numerous upstream unit operations, and the EAF. Moreover, an entire year of high frequency operating data is collected and analyzed to develop the arc loss predictor.

The goal of the fault predictor is to provide operators with a warning five to ten minutes in advance of an event with a 75% or higher probability of inducing arc loss such that operators can take preventative measures. Operators require at least two minutes prior to the arc loss event in order to take the neccesary corrective actions. Figure 3.3 illustrates the entire ML workflow including the data preprocessing tasks resulting in segmented datasets ready for representation learning and predictive classification. Note, some classification methods bypass explicit representation learning and instead learn from the raw features. There are also hidden feedback connections between the modules as the workflow progresses in a largely iterative fashion. Initially this alarm will serve as a tool for engineers and operators

but ultimately the goal is to implement an advanced controller that can automatically take corrective action.

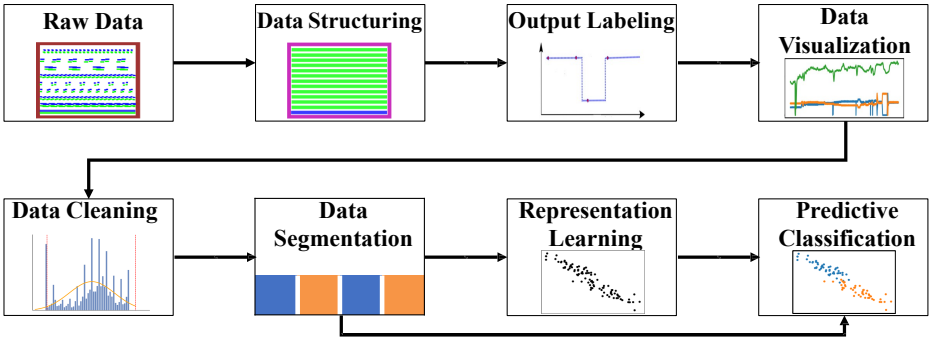


Figure 3.3: Flowchart illustrating the overall data analytics workflow.

The novel contributions presented in this work include the introduction of the DC EAF arc loss FDD challenge and the formulation of this fault as a supervised ML problem. Successful problem formulation is a significant contribution that includes transforming a year of raw industrial operating data into cleaned, structured, labeled, and segmented datasets that are amenable to further statistical ML analysis. Labeling the data requires the introduction of rigorous quantitative conditions to detect the arc loss. Given a precise problem formulation and procured training data, the remaining contribution is the development of the arc loss prediction inferential sensor. This contribution also includes a comprehensive validation and comparison of traditional and advanced approaches to representation learning and predictive classification on industrial operating data.

3.2 Data Preparation and Visualization

This section presents a comprehensive account of the data preparation tasks between obtaining raw industrial process data, formulating a supervised learning problem, preprocessing the data for ML, and data visualization. Problem formulation and data preparation present endless opportunities for improvement. This case study serves as an important end-to-end experience that demonstrates the full stack of activities involved with developing a soft sensor from industrial data to address

an important process fault. The subsequent case study on the kraft pulping process provides less detail on the fundamental data preparation tasks and more emphasis on process domain knowledge.[2]

Data preprocessing produces the datasets that are used to train, validate, and test the predictive models. Data preparation procedures have significant consequences on the generalization performance of a supervised ML algorithm [61]. The goal of preprocessing is to transform the raw historian data into a form that is amenable for statistical ML algorithms. In this case study, the objectives of data preparation include maximally retainining information from the raw data, minimizing extraneous information injected during preprocessing, and removing redundant data. This section includes a description of the raw data followed by an overview of the methods used to structure, visualize, clean, and segment the data.

3.2.1 The arc loss dataset

The raw data used in this work involves one year of daily exports from an industrial process historian. The scope encompasses relevant operating data collected in 2017 from a large-scale integrated open-pit mining and metallurgical process. Each day of operation is captured and stored as a comma separated value (CSV) file with approximately 228 columns and thirty thousand rows. Half of the 228 columns are measured variables and the other half are corresponding timestamps. There are 92 PVs with high frequency measurements (e.g., feed rate, temperature, etc.), 14 variables with label encoded discrete values (e.g., valve open or closed), and five laboratory measurements with sample periods greater than one hour. Overall, the total daily exports have an uncompressed size of 17.4 gigabytes which can make the data unwieldy for practitioners to load, analyze, and process. Detailed descriptions of the data are provided in Table A.1 of Appendix A.

3.2.2 Data structuring and output labeling

The columns of each daily export have a varying number of rows with more densely sampled PVs having up to thirty thousand rows and others having as few as ten

[2]This is in part due to necessity, but also due to data access limitations with the pyrometallurgy study.

samples. The raw data contains asynchronous data with both numerical PVs such as furnace temperature and categorical variables such as valve positions. The raw data contains errors such as missing values, bad inputs and not a number (NAN) values. Systematically structuring the raw data and removing the corrupted data is one of the first stages of data preparation.

Each of the 365 daily CSV file exports is processed to replace non-numeric inputs (e.g., 'tag not found') with NAN values and remove rows and columns with overwhelming NAN values (e.g., rows with less than three non NAN values). The illustration in Figure 3.4 shows the structuring of three consecutive days and represents time horizontally. Each operating variable is represented by a green row and the accompanying timestamp is represented by a blue row with the dashes representing the differing frequency of measurements. The top of Figure 3.4 shows the structured dataset that has one unified timestamp and no NAN values. To preserve information the time index corresponding to the most densely sampled variable is used as a unified timestamp (blue bars at the top of Figure 3.4). To minimize insertion of synthetic data the less frequently sampled variables are re-sampled using a simple forward fill or zero order hold operation.

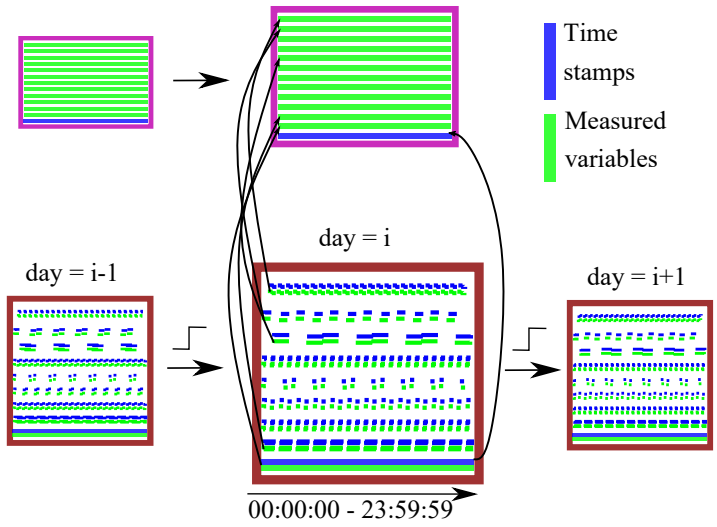


Figure 3.4: Structuring consecutive days of historical data.

Once the data is cleaned and structured it is now suitable for generating the arc loss labels. The open plasma arc is developed when the electric current between the graphite electrodes (i.e., cathode) and the surface of the molten slag (i.e., anode) is charged. During routine operation, the power sent to the furnace is relatively stable and the electrodes are held in a stationary position. However, due to process disturbances (e.g., varying slag composition, upstream disturbances in furnace feed, and electrical noise from the power supply) the power applied to each graphite electrode can vary and arc loss can occur.

The power measured at each electrode can be monitored to detect arc losses. To ensure the labels are robust, the arc loss problem is defined when all three of the conditions illustrated in Figure 3.5 are satisfied. The conditions are as follows:

1. The power is otherwise stable (i.e., the standard deviation of the power applied to an electrode is less than 2 MW over a period of approximately 11.5 minutes).

2. A precipitous power drop (i.e., at least 10 MW within thirty six seconds).

3. The power recovers (i.e., to within 5 MW of the original stable value within approximately 10 minutes).

These three conditions are rigorously applied to each sample for both electrodes to generate output labels that are binary indicators of arc loss in the respective electrode. Accurately labeling the dataset is critical for supervised learning.

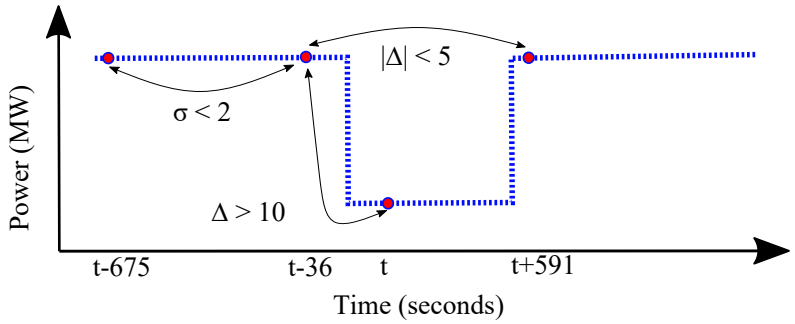


Figure 3.5: Illustration and quantitative definition of conditions constituting arc loss.

3.2.3 Data visualization

Data visualization provides key insights into the frequency of the faults and the severity of the arc loss on operating efficiency. Visualization also assists in troubleshooting, validation of data pre-processing, and output labeling. An integrity check is performed on the binary arc loss labels to confirm that they correspond to a representative power drop. Figure 3.6 shows three discrete arc loss labels in the top plot and the corresponding arc 'A' power upsets[3] in the bottom plot.

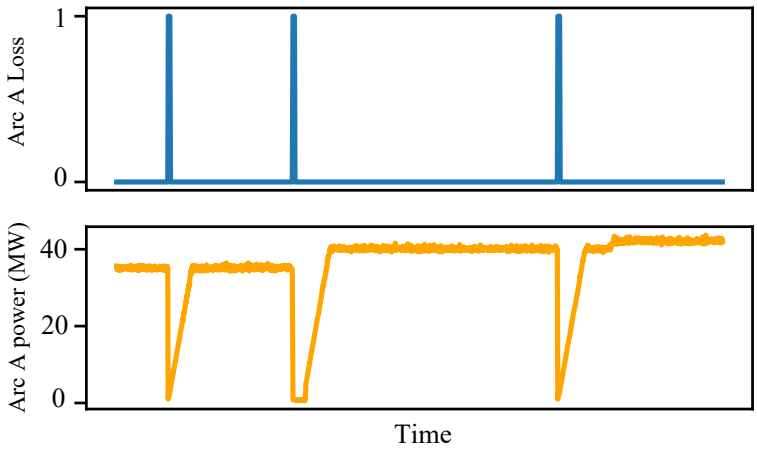


Figure 3.6: Visual validation of the arc loss labels.

The frequency of arc loss events is clear from Figure 3.7 which shows the number of arc loss events per day for each plasma arc throughout a year of operation. Arc loss is a significant problem that can occur as often as twenty five times per day (indicating a chain of arc losses) or not at all for multiple consecutive days. This distinction provides motivation to apply data-driven pattern recognition techniques to determine the difference in operation between arc loss cascades and stable operation. Although the average duration of an arc loss label is less than one minute, the disruption to the EAF of a single loss event can cause up to twenty minutes of lost production. This visualization not only provides motivation but it also helps to recognize the class imbalance in our output labels due to the short average duration

[3]The EAF has two graphite electrodes ('A' and 'B') that generate plasma arcs.

of each arc loss indication.[4]

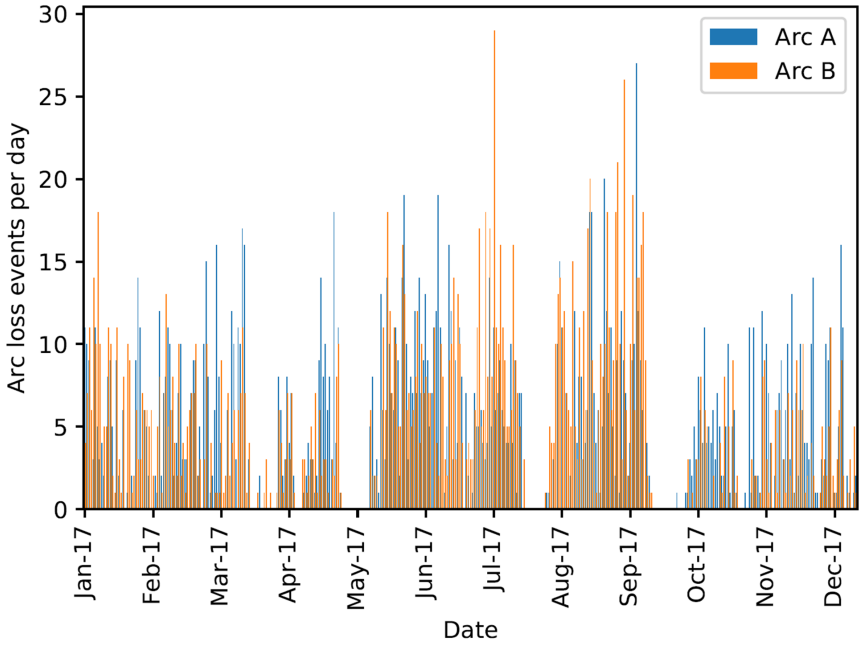


Figure 3.7: Daily arc loss events in each electrode over one year of operation.

Finally, the severity of the arc loss fault on EAF operation is visualized in Figure 3.8 by comparing a period of relatively stable operation (top) to a period of faulty operation (bottom) using the power applied to each electrode and the furnace feed rate. The arc loss fault has a significant impact on the furnace feed rates and subsequently on the production rate of the EAF which can bottle-neck the entire metallurgical operation. Thus, it is imperative to prevent loss of the plasma arc in order to sustain economic viability of the process.

3.2.4 Data cleaning

The quality of any ML model depends on the quality of the input it receives. Here, data cleaning involves setting PV limits using process knowledge to filter out nonsensical values (e.g., negative feed rates), removing problematic PVs, and removing

[4]Class imbalance is an important consideration addressed at the end of this section.

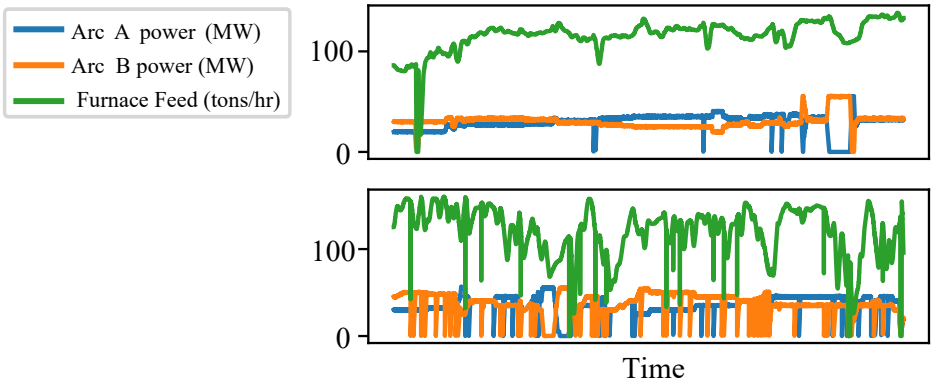


Figure 3.8: Visually comparing relatively stable operation (top) to faulty operation (bottom) to observe the impact of arc loss upsets [108].

data from plant shut-downs. Erroneous process data and outliers can induce spurious correlations and increase the rate of misclassification for ML classifiers [3]. Removal of this data is accomplished through domain expertise and consultations with industrial collaborators. A set of minimum and maximum limits are agreed upon for each PV and measurements outside of these limits are set to either the nearest limit or three standard deviations from the mean.

The left side of Figure 3.9 shows the power values for arc 'A' as a histogram with a normally distributed probability density function (PDF) and PV limits shown by the vertical red lines. There are some negative power values that are subsequently adjusted to zero during data cleaning. Using process knowledge to set PV limits is not an infallible strategy. Visual verification is often necessary to ensure the PV limits are correct as demonstrated by the right side of Figure 3.9 which shows the crucible heat loss as a PDF with the original PV limits as vertical red lines. All of the crucible heat loss data is outside the original PV limits, but instead of cleaning this data the PV limits are re-evaluated and it is deemed acceptable. This PV limit verification procedure is conducted for all of the PVs.

Outliers are often considered to be values that are greater than three standard deviations from the mean. Box plots are commonly used to show the distribution of a variable and indicate the number of outliers. Treatment of outliers is application

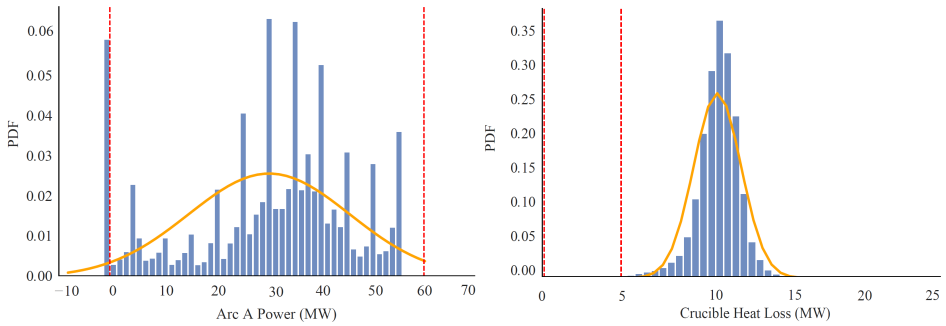


Figure 3.9: Setting PV limits with process knowledge to remove invalid data.

specific and modeling abnormal behavior requires retention of data that may be statistically defined as outliers. Simple first-principles are used to remove invalid outliers (e.g., negative furnace feed rates).

Two final tasks remain for data cleaning, i.e., removing unhelpful PVs and removing irrelevant shutdown data. Five laboratory measurements are deemed to have too low of a sampling frequency for use as a fault predictor and were therefore removed entirely from the dataset. Seven PVs were removed based on prior knowledge of having unreliable, faulty or inaccurate measurements. Two plant shutdowns are clearly visible in Figure 3.7 in May and early October. The data for these periods is carefully removed for all PVs during data cleaning to preserve useful information during startup and shutdown phases. By carefully structuring and cleaning the data a significantly smaller set of data is created that preserves useful process information and is more amenable to subsequent modeling.

Data segmentation

For binary classification problems with a large degree of class imbalance, the vast majority of instances fall into the majority class while significantly fewer instances fall into the minority class (i.e., the class of interest for fault detection). Most binary classification methods perform poorly on imbalanced datasets due to assuming the data are drawn from the same distribution and assigning equal weight to both classes. Classifiers aim to achieve the highest accuracy along the whole range of data and therefore tend to largely ignore the minority class which has relatively

negligible impact [136].

Previous studies suggest techniques to address class imbalance can be divided mainly into three categories: re-sampling, feature engineering, and classifier manipulation [64]. Artificially re-sampling the instances to balance class distributions can be performed by either under-sampling the majority class or over-sampling the minority class [78]. Under-sampling is at risk of discarding information from the majority class while over-sampling increases the likelihood of over-fitting by duplicating instances from the minority class [41]. More advanced methods include explicitly combining separate features from the minority and majority classes as well as manipulating the classifier weights internally [156].

The arc loss dataset is highly imbalanced with 99.67% of the samples labeled as the majority class (i.e., no arc loss) and only 0.33% of the instances labeled with arc loss. An under-sampling approach is taken to address class imbalance by extracting a segment that contains 55 minutes worth of data in the 5-60 minute period before every arc loss. All 1526 arc loss events (taken from both arc A and arc B) are extracted to represent the minority class. The majority class is randomly under-sampled and only 1526 segments that correspond to 55 consecutive minutes taken 5 minutes prior to periods of extended stable operation are extracted.

The data segmentation process is illustrated in Figure 3.10. The entire dataset, containing 3052 segments, is further divided with 85% (or 2594 balanced segments) for cross-validated training and 15% (or 458 balanced segments) for testing. With the data finally procured to a suitable format it can be used to train the representation learning and predictive classification algorithms.

3.3 Learned Representations and Predictive Classifiers

Representation learning is described as a subset of ML and a superset of deep learning. Classical ML is distinguished from representation learning through the selection of features. In classical ML features are hand-designed whereas in representation learning features are learned from the data. Moreover, in deep representation learning there are numerous layers of abstraction between simple learned features and more complicated features that may further improve representation [38].

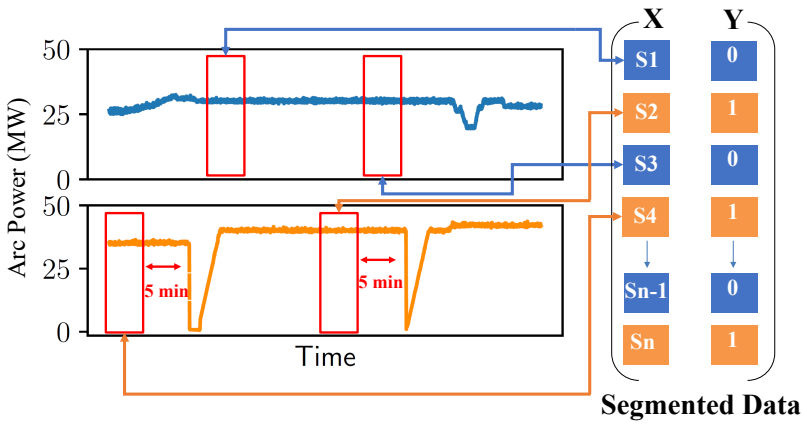


Figure 3.10: Illustration of data segmentation to create a balanced dataset.

Representation learning is defined as the means by which an efficient and informative representation can be learned that extracts useful information to improve the performance of classification, regression or prediction models [10]. The labor intensive procedure of engineering features is excellent at leveraging application specific domain knowledge but it lacks efficiency and ease of applicability across various domains [137]. Representation learning can learn the important discriminatory features from the data in a systematic fashion allowing for faster deployment of AI in a variety of domains [10].

The use of deep representation learning algorithms for PSE applications (e.g., control, process monitoring, and fault detection) is a relatively new, but highly active research area [43]. The TEP has been used as an FDD benchmark to validate many advanced neural architectures including stacked sparse AEs [81], deep belief networks [143], and deep CNNs [142]. These studies are limited in their demonstrations with industrial case studies and their focus on fault detection, as opposed to fault prediction. This work validates and compares traditional and advanced process monitoring methods on historical data taken from a large-scale industrial process with the challenging goal of predicting arc loss five minutes before it occurs.

This work focuses on studying and validating the benefits of using representation learning (e.g., dimensionality reduction) and deep learning for predictive classifica-

tion with industrial operating data. Selected methods are illustrated in Figure 3.11 with explicit representation learning algorithms on the left and the predictive classifiers on the right. Explicit representations are learned using PLS and PCA; while logistic regression (LR), linear support vector classifiers (L-SVCs), kernel support vector classifiers (K-SVCs), ANNs, and CNNs are all compared as predictive classification models. Note, the not applicable (N/A) indicates the use of raw features instead of an explicit representation. Also, the K-SVC, ANN, and CNN methods have internal representations with kernels, hidden layers and convolutions, respectively. Altogether, Figure 3.11 shows fifteen experimental combinations with seven algorithms that are introduced in what follows.

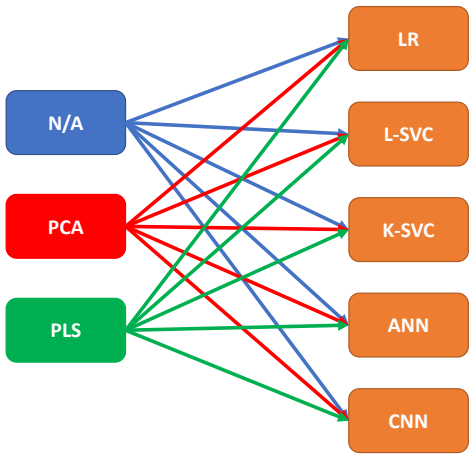


Figure 3.11: Experimental configurations with representations (left) and predictive classifiers (right).

3.3.1 Explicit representations with reduced dimensionality

Two traditional process analytics methods (i.e., PCA and PLS) are applied to learn explicit dimensionally reduced representations from raw features. Both are popular FDD methods, but they differ notably in that PCA is an unsupervised learning algorithm, and PLS is a supervised learning algorithm. These methods were both introduced in Chapter 2 as quantitative process history based methods for FDD.

3.3.2 Predictive classification and implicit representations

The right side of Figure 3.11 lists the five predictive classification methods that are trained and tested with either the raw data or the explicit representations learned through PCA and PLS. These predictive classifiers are introduced and discussed briefly in what follows to provide experimental background information.

Logistic regression

Choosing LR for binary classification is natural as the standard logistic function (i.e., the sigmoid function) given by

$$P(Z) = \frac{\exp(Z)}{1+\exp(Z)} = \frac{1}{1+\exp(-Z)} \tag{3.1}$$

provides a bounded output between zero and one that can be interpreted as the probability of a binary outcome and mapped to discrete classes (e.g., arc loss or no arc loss). The input, $Z = \alpha + \beta X$, to the logistic function illustrates the connection with linear regression where X is the pre-processed data (or a learned representation thereof), α is a scalar bias, and β is a weight vector. Historically, LR dates back to the early 19th century when the logistic function was invented to describe population growth and autocatalytic chemical reactions [26]. Recent applications of LR in PSE include methods that combine LR with dominant trend extraction and dependent binary relevance classifiers to perform nonstationary fault diagnosis and multi-label fault classification, respectively [100, 114].

Support vector classifiers

The basis for the L-SVC predictive classification technique used in this work is the soft margin support vector machine (SVM) introduced in 1995 which is itself an extension of the hard margin SVM, conceptually solved in 1965 [25, 131]. The difference between hard margin and soft margin SVM is that hard margin SVM assumes the classes are linearly separable and thus tries to find a hyperplane such that no point is misclassified, whereas soft margin SVM allows for some misclassification that is proportionally penalized in the objective function. Binary support vector classifier (SVC) aims to construct a separating hyperplane between the two

classes of data such that the margin (i.e., distance) between the hyper-plane and the nearest data points of each class is maximized [140].

Nonlinear formulations of SVMs utilize the kernel trick, i.e., K-SVC, such as the parametric polynomial kernel or the non-parametric radial basis function (RBF) kernel with important properties that allow for enhanced representation capacity and efficient optimization [17, 113]. Recent applications of SVMs in PSE include applying one-class SVM on finite impulse response (FIR) data to detect MPM in a paper machine control system [79, 105] and using nonlinear SVM-based feature selection for FDD [95]. This work studies linear and kernel based SVCs with a variety of configurations (e.g., kernel and regularizer choices) described in Section 3.4.

Artificial neural networks

A brief introduction to ANNs was provided in Chapter 2 while discussing deep learning for fault detection. In the context of binary classification, the output layer of the ANN consists of a single output neuron that indicates the class of the segment by computing the weighted sum of hidden values from the last hidden layer, followed by a sigmoid function, i.e.,

$$Z = (\sum_{i=1}^{n} x_i^L w_i^L + b) \qquad y = f(Z) = \frac{\exp(Z)}{1+\exp(Z)}, \tag{3.2}$$

where x, w, and b are as described in equation 2.2. The superscript L refers to values and weights from the neurons in the final hidden layer. If the output of the sigmoid neuron is greater than or equal to 0.5, it outputs 1 (i.e., arc loss). However, if the output is less than 0.5, it outputs 0 (i.e., stable operation). In this work, flattened input segments are fed to a multi-layered fully connected perceptron model to predict arc loss.

Convolutional neural networks

The CNN architecture is introduced in Chapter 2 along with key concepts such as convolutional layers, learnable kernals, and feature maps that are not re-introduced here. Applications of CNNs in PSE include FDD on the TEP [142], a three phase

flow facility at Cranfield University [151], and in a semiconductor manufacturing process [73]. To my knowledge, this work represents the first time CNNs have been studied for fault prediction in an industrial manufacturing process with historical operating data [108].

3.4 Experimental Setup

The experiments primarily consist of training, validating, and testing the fifteen experimental configurations shown in Figure 3.11. The experimental setup has the following two key factors that distinguish this work from previous FDD studies in PSE: i) simulating a production trial by preserving the temporal integrity of our training data with respect to our testing data, and ii) performing rigorous cross-validation and hyperparameter optimization to compare models.

The preprocessed data segments are split into two groups; the training and validation group consists of 2594 segments and the testing group consists of 458 segments. Prior experimental designs performed random selection of segments for training and testing sets throughout the entire year of operation. Random sampling is common in literature as well, but because the objective is to develop an inferential sensor for an industrial process our experimental design mimics that of a production trial. As shown in Figure 3.12, the simulated production trial trains and validates models on the first ten months of operation while the last two months of operation are strictly used for testing the final models.

3.4.1 Stratified k-fold cross-validation

A stratified k-fold cross-validation strategy is used to compare different hyperparameter configurations in the predictive classifiers. As shown in Figure 3.13, ten non-overlapping folds are created where each fold contains a balanced number of arc loss segments and stable operating segments. For each hyperparameter optimization trial (i.e., corresponding to a specific configuration) the model is trained on 90% of the training data and validated on the remaining 10%. This is repeated ten times, once for each fold, where the validation data changes as shown by the yellow highlight in 3.13. The result of the trial is the average accuracy of all ten

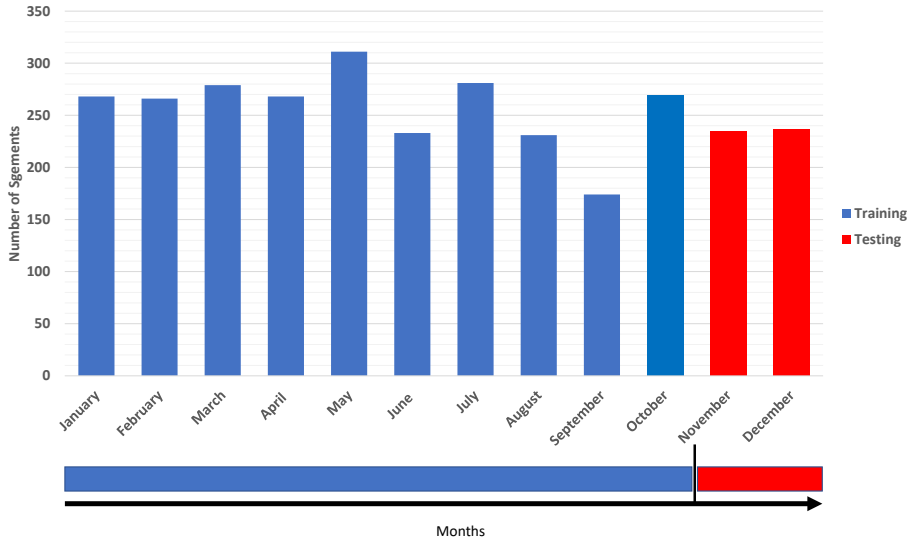


Figure 3.12: Splitting the data into training and testing sets based on date.

validations which serves as a score to rank the hyperparameter configuration.

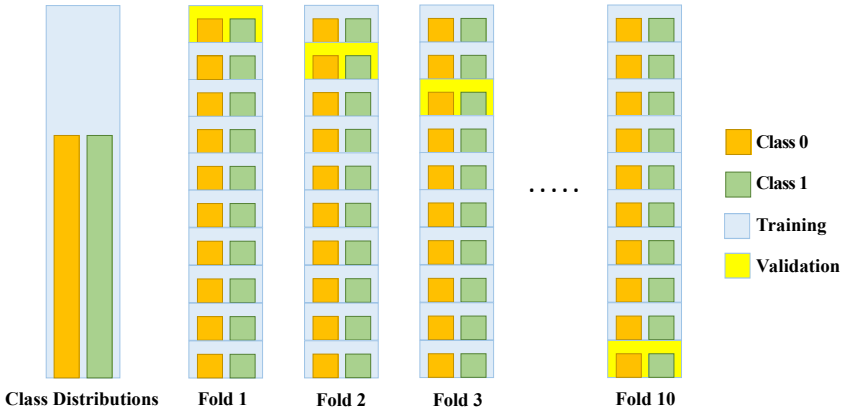


Figure 3.13: Each experimental configuration is trained and validated with stratified k-fold cross-validation [108].

3.4.2 Hyperparameter optimization

A robust and transparent hyperparameter optimization strategy is critical for an impartial comparison of ML algorithms and the reliable development of a predictive inferential sensor. The efforts taken here aim to contribute a high level of rigor to hyperparameter optimization in the context of PSE. A broad space of possible hyperparameters is defined for each predictive classifier and then a Bayesian sequential model-based optimization (SMBO) algorithm searches this space using a tree-structured Parzen estimator (TPE) to suggest the best configurations by maximizing expected improvement [13][12]. Multiple trials are conducted for each hyperparameter configuration where the TPE specifies the configuration for the next trial based on the expected improvement.

Optimizing hyperparameters for learned representations

Recall from Figure 3.11, the predictive classifiers are provided segments in the raw feature space or a latent space of reduced dimensionality using either PCA or PLS. The only hyperparameter that is considered while generating the PCA and PLS representations is the number of components for each method, i.e., the dimensionality of the latent space. Exhaustive search is performed by performing ten-fold cross-validation with LR classification while iteratively increasing the number of components for PCA and PLS. Figure 3.14 shows the resulting validation accuracy as the number of components increases. The peak validation accuracy occurs at 16 components for PLS (i.e., PLS-16) and 41 components for PCA (i.e., PCA-41). Each of the five predictive classifiers is optimized and tested with data from three representations, i.e., raw features, PCA-41, and PLS-16.

Optimizing hyperparameters for predictive classifiers

The Bayesian SMBO is performed on the predictive classifiers. The number of trials for each method is manually selected based on the size of the hyperparameter search space along with consideration for computational limitations. Optimizing the hyperparameters for these models, particularly for the deep learning models, is by far the most computationally demanding aspect of this work.

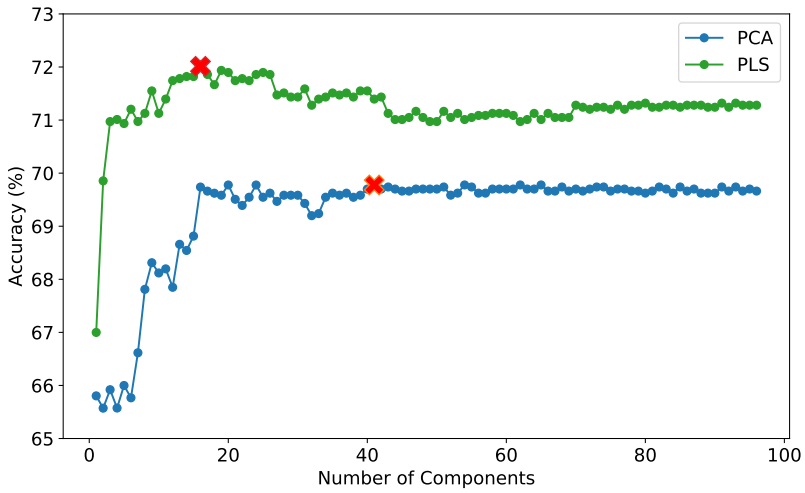


Figure 3.14: Selecting the number of components for PCA and PLS representations with cross-validation. Note that a higher accuracy is obtained with the supervised PLS algorithm.

The classifiers are divided into traditional ML algorithms (i.e., LR, L-SVC, and K-SVC) and deep learning algorithms (i.e., ANNs and CNNs) for the purpose of describing the hyperparameter optimizations. The hyperparameter search space for the traditional ML algorithms and the deep learning algorithms is provided in Table A.2 and Table A.3 of Appendix A, respectively. Search options for the tolerance and the regularizer strength (λ) are the same for all traditional methods but some of the remaining hyperparameters only apply to one method (e.g., kernel type only applies to K-SVC). Notably, the penalty denoted elast. refers to elastic net, the squared-hinge (SH) loss is abbreviated, and the three kernels are abbreviated as poly. for polynomial basis, sig. for sigmoid basis, and the RBF. A thorough explanation of each hyperparameter is beyond the scope of this work. Inquisitive readers are referred to literature and documentation for further information [97].

For each traditional ML classifier there are three explicit representations of the data that are separately optimized for hyperparameters. The result is nine experimental configurations of traditional ML algorithms with optimized hyperparameters spec-

ified in Table 3.1. Notably, an RBF kernel was selected for all K-SVCs, an SH loss was selected for all L-SVCs, and the optimized regularization strength varies significantly depending on the representation.

Table 3.1: Hyperparameter selection for traditional ML algorithms.

		λ	tolerance	penalty	loss	kernel
NA	LR	10	0.001	elast.		
	L-SVC	1000	0.001	ℓ_1	SH	
	K-SVC	0.1	0.0001			RBF
PCA-41	LR	10	0.001	elast.		
	L-SVC	100	0.00001	ℓ_1	SH	
	K-SVC	0.01	0.0001			RBF
PLS-16	LR	0.001	1e$-$5	ℓ_2		
	L-SVC	100	0.001	ℓ_1	SH	
	K-SVC	0.1	1e$-$5			RBF

Although more hyperparameter optimization trials are conducted for the deep learning methods, the percentage of the search space covered by these trials is significantly smaller. This is due to the fact that the search space for the deep learning methods is orders of magnitude larger than the traditional ML methods. It is infeasible to conduct enough trials to search over an equivalent percentage of such a large space. Even with the use of cloud computation platforms to mitigate computational limitations, there are practical limitations on the number of convolutional layers, batch size, and the number of learned filters in the studied architectures.

After a series of challenging cross-validation trials the final choice of ANN and CNN hyperparameters are shown in Table 3.2 for each of the representations. The choice of optimizers, regularization strengths (λ), and fully connected layer (FCL) activation functions are the same for both ANN and CNN models. Some hyperparameters that are unique to the CNN include the number of convolutional layers (CLs) and the number of learnable filters.

Ultimately, the fifteen experimental configurations in Figure 3.11 are outfitted with the parameters in Table 3.1 and Table 3.2. These models are tested on segments

Table 3.2: Hyperparameter selection for deep learning algorithms.

representation	NA		PCA-41		PLS-16	
classifier	ANN	CNN	ANN	CNN	ANN	CNN
optimizer	Adagrad	SGD	Adam	SGD	Adagrad	RMSprop
λ	0.1	0.05	0.01	0.001	0.1	0.0001
FCL activation	relu	elu	elu	tanh	elu	elu
no. of FCLs	5	2	9	1	10	1
FCL size	128	32	32	24	128	64
batch size	128	32	128	16	32	32
no. of CLs		1		1		1
CL activation		tanh		elu		tanh
filters		6		24		16
filter size		(5,5)		(20,20)		(3,3)
pool size		(1,1)		(2,2)		(2,2)

from two months of subsequent operation as described in what follows.

3.5 Results and Discussion of Arc Loss Prediction

This section presents the experimental results of testing the arc loss inferential sensor. Discussion is also provided that summarizes the contributions, analyzes findings, and suggests improvements.

3.5.1 Experimental results

Classification results from supervised learning studies can be represented as a contingency table, known as a confusion matrix, with a dimension for true class values (i.e., $y = 1$ or $y = 0$) and a dimension for predicted class values (i.e., $\hat{y} = 1$ or $\hat{y} = 0$). An arc loss is considered a positive event ($y = 1$), and no arc loss (i.e., stable operation) is considered a negative event ($y = 0$). The confusion matrix consists of four values; two of which correspond to correct predictions, and two of which correspond to false predictions. False predictions can be either false positive (FP) or false negative (FN), referring to either type I error (false alarm) or type II error (missed alarm), respectively. True predictions can be either true positive (TP) or

true negative (TN), i.e., correctly predicting arc loss or correctly predicting no arc loss, respectively.

For a particular experimental configuration (e.g., LR with PCA), the model produces an output estimate for each segment in the testing set. Each output estimate is compared to the true output label allowing the categorization of that prediction as either FP, FN, TP, or TN. Therefore, the sum of these four values is equivalent to the total number of segments in the testing dataset and the resulting confusion matrix summarizes the prediction fidelity of the model with respect to both the positive and the negative class. Various performance metrics (e.g., accuracy) can be derived for each model from the confusion matrix of that model.

The confusion matrix resulting from testing each of the fifteen experimental configurations in Figure 3.11, with parameters shown in Table 3.1 and Table 3.2, is provided in Table 3.3. In addition to the confusion matrix, two key performance

Table 3.3: Summary of the experimental results.

		TP	FP	TN	FN	ACC	TPR
NA	LR	182	49	151	76	0.727	**0.705**
	L-SVC	185	46	141	86	0.712	0.683
	K-SVC	167	64	156	71	0.705	0.702
	ANN	176	55	152	75	0.716	0.701
	CNN	181	50	141	86	0.703	0.678
PCA $d = 41$	LR	186	45	148	79	**0.729**	0.702
	L-SVC	176	55	151	76	0.714	0.698
	K-SVC	144	87	158	69	0.659	0.676
	ANN	176	55	142	85	0.694	0.674
	CNN	181	50	149	78	0.721	0.699
PLS $d = 16$	LR	184	47	130	97	0.686	0.655
	L-SVC	169	62	146	81	0.688	0.676
	K-SVC	166	65	149	78	0.688	0.680
	ANN	166	65	147	80	0.683	0.675
	CNN	147	84	161	66	0.672	0.690

metrics are tabulated in Table 3.3 for each configuration, i.e., the accuracy (ACC)

and the recall, otherwise known as the true positive rate (TPR), with maximum values emphasized in bold font. Accuracy is simply defined as the sum of true predictions (i.e., TP and TN) divided by the sum of all predictions (i.e., the total number of segments). The most accurate experimental configuration is with an LR classifier on a 41 principal component representation followed very closely by an LR classifier on the raw data itself.

The second critical performance metric provided in Table 3.3 is the recall which focuses on the cases which precede an arc loss event. Specifically, recall is defined as the number of times arc loss is correctly predicted divided by the number of times arc loss occurs, i.e., $\mathrm{TPR} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN})$. The experimental configuration with the best recall in this study is a logistic regression classifier on the raw data. Interestingly, the runner-up for recall is a tie between a logistic regression classifier on a 41 principal component representation and a kernel support vector classifer on the raw data.

Aside from accuracy and recall, performance metrics for precision, also known as positive predictive value (PPV), F_1 score, and F_β score are tabulated in Table A.3 in Appendix A. The F_1 score represents the harmonic mean of precision and recall, whereas the F_β score allows user specification of β which controls the weighting of recall and precision (i.e., recall is β times more important than precision). For this application, recall is prioritized because the operating cost associated with false alarms is much less than the operating cost associated with missed alarms (i.e., FN). Therefore, a choice of $\beta = 0.25$ is selected for tabulating the F_β scores. The best configuration with respect to precision and F_1 score is an LR classifier with a PCA representation, whereas the configuration with the highest F_β score is, unsurprisingly, the same as that with the highest recall, i.e., an LR classifier on the raw data.

Comparing representations in terms of accuracy, the raw data and the PCA representation consistently outperform the PLS representation with the only exception being for the K-SVC classifier. Comparing classifiers in terms of accuracy, the LR classifier has the two highest accuracy scores with 72.9% and 72.7% on PCA components and raw data, respectively.

With respect to the comparison of recall scores across the different representations, i.e., PLS is once again outperformed by PCA and raw data. Simply using the raw data provides the best recall score on average (across classifiers) as well as the highest recall of 70.5% with an LR classifier. Overall, logistic regression demonstrates better generalization performance relative to the deep learning methods. Ultimately, given the importance of recall in this application, a logistic regression classifier on the raw data is the most promising configuration for development of an inferential sensor to predict arc loss.

Deep learning methods contain a very large number of parameters which allows them to model complex nonlinear functions if they have enough data to train on. Although the logistic regression method performed slightly better in these experiments, it is possible that the deep learning methods would perform better in an experiment with multiple years of historical operating data. This is demonstrated by Figure 3.15 which shows the superior performance increases with deep learning methods relative to traditional methods when additional data is provided. Another

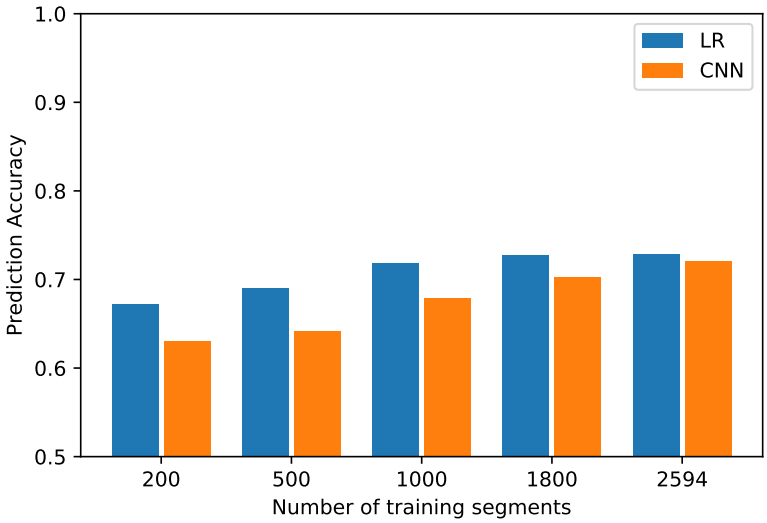


Figure 3.15: Comparing the classification accuracy of logistic regression with convolutional neural networks while varying the amount of training data.

common issue with deep learning methods is over-fitting. Special care is taken to prevent over-fitting by introducing regularization and early stopping. However, over-fitting is still a significant challenge for deep learning applications with soft sensors. Finally, sensitivity to hyper-parameters and network initialization is another potential concern for deep learning methods.

3.5.2 Discussion of contributions and findings

This case study has introduced a novel industrial predictive classification problem, i.e., to predict arc loss in a DC EAF five minutes prior to the arc loss event. Moreover, an end-to-end ML workflow is presented that demonstrates how raw industrial data can be used to deliver operating value in the form of an inferential sensor model. Arc loss events are predicted five minutes prior to occurrence with an accuracy of 72.7% and a TPR of 70.5% on unseen data from two subsequent months of operation. The unexpected loss of plasma arc in the DC EAF is an ongoing problem resulting in millions of dollars of lost production annually. This work has the potential to contribute improved economic savings and better environmental outcomes as energy and material are consumed more efficiently. Finally, a comprehensive empirical comparison between traditional and contemporary ML methods is presented for both representation learning and predictive classification.

Given the recent success of deep learning methods it is interesting to note that applying the logistic regression classifier to the raw data is the best performing experimental configuration. Especially considering a significantly greater computational effort is taken to optimize and train the deep learning methods relative to the traditional ML classifiers. It is important to consider whether or not this optimization effort contributed to over-fitting and reduced model generalization.

Concerns for over-fitting are also present in the study of explicit representations where the raw data provides superior test results. The poor performance of PLS relative to PCA is consistent with over-fitting given that PLS is a supervised learning method with an iterative optimization procedure. Note that Figure 3.14 demonstrates a consistently higher validation accuracy for PLS relative to PCA.

Increasing the fidelity of the experimental setup with respect to the constraints of

the production environment is a non-trivial challenge that is critical for improving the value of data-driven insights. For example, Figure 3.15 suggests that increasing the amount of data used to train the deep learning models may help overcome overfitting and deliver enhanced generalization. The experiments can also be enhanced by evaluating the algorithms with continuous raw process data instead of procured class-balanced segments. A change that would likely improve performance while maintaining operational integrity is to decrease the period over which the model is evaluated without being updated. The predictive model could be updated more frequently than every two months (e.g., weekly) to improve performance. Finally, ongoing work includes the development of a benchmark challenge with the arc loss dataset that can supplement existing simulated benchmarks (e.g., TEP).

3.6 The Arc Loss Challenge

As the kraft pulping case study demonstrates, the ability to conveniently and accurately label arc loss is a key differentiating characteristic of the pyrometallurgy study. As Chapter 2 describes, open-source benchmark datasets have great potential to assist researchers in developing and comparing ML methods for FDD. Literature review on the TEP benchmark helped direct the method selection in the representation learning study. This review also helped reveal limitations of existing FDD benchmarks and identify an important opportunity to address these limitations by publishing a modern benchmark challenge with raw operating data from a large-scale industrial process, i.e., *The Arc Loss Challenge.*

Framing the unexpected loss of plasma arc in an electric arc furnace as a supervised learning problem is a novel contribution to process monitoring literature. Development of a soft sensor with readily available operating data is demonstrated with 72.9% accuracy on a testing data set. Further improvements are recommended before this solution is adopted in practice. Arc loss forecasting is an open problem that can help operators take preventative measures to minimize the negative impact of arc loss on production. *The Arc Loss Challenge* provides the industrial operating data with arc loss labels as a structured FDD benchmark.

The long-term utility of ML benchmarks in many disciplines is sensitive to the con-

cept of metric fixation which is succinctly described by Goodhart's law, i.e., "*when a measure becomes a target, it ceases to be a good measure*" [121]. Concentrating the measure of progress for an entire discipline on the results of a single benchmark is a form of high order over-fitting. This scenario is even more myopic when the discipline is over-fitting to a benchmark with highly procured simulated data that does not represent many of the non-trivial challenges encountered in practice (e.g., data preparation). For instance, the most prominent FDD benchmark is the TEP dataset which is used in literature with increasingly exceptional results (i.e., accuracy in the range of 96% to 100%) using advanced ML algorithms [145].

One way to reduce the illusion of progress caused by this high-order over-fitting is to maximize the alignment of the challenges posed by the benchmark with those faced in practice. Another remedy is to encourage benchmarking that simultaneously helps mitigate unresolved process faults that continue to burden real-world industrial operations. Finally, the utility of benchmarking strategies can be improved by posing non-trivial challenges for modern ML methods, ensuring consistent usage of the benchmark, and encouraging transparency so proposed methods can be replicated and examined.

The Arc Loss Challenge is introduced to help remedy the limitations of existing ML benchmarks for FDD. By providing an open-source dataset from a large-scale industrial process with accurate labels of an ongoing process fault, *The Arc Loss Challenge* helps researchers measure progress that aligns with real operating environments. A clear and concise problem formulation is published with the data along with consistent and transparent procedures for submission and evaluation.

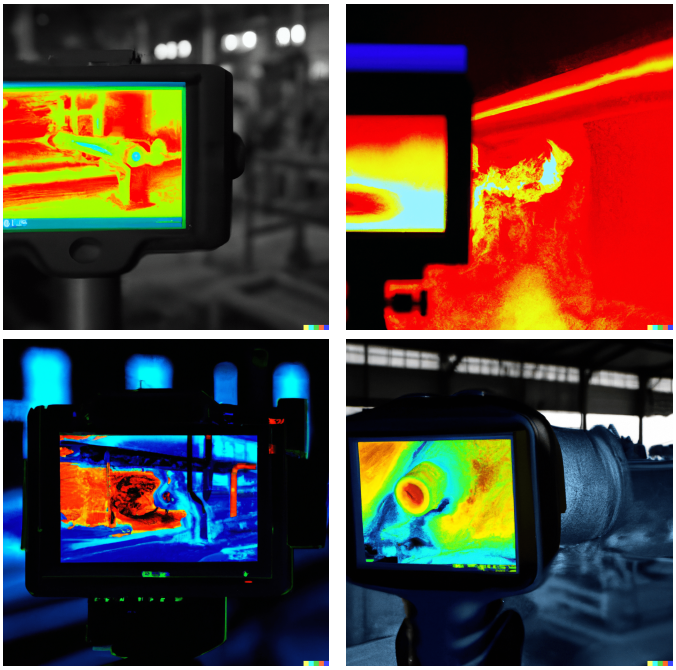
As mentioned before, *The Arc Loss Challenge* has been accepted for publication in the *The Journal of Process Control*. By providing a supervised learning challenge based on large quantities of raw industrial process data with transparent and consistent evaluation procedures, *The Arc Loss Challenge* is a unique contribution to fault detection benchmarking. To participate in *The Arc Loss Challenge*, interested readers are directed to the competition website [148].

In this chapter, minimal process knowledge is used to introduce the arc loss fault and develop an ML solution. Chapter 4 presents a fault that does not benefit from

such a distinct definition, and a dataset that is not as easy to accurately label. A higher degree of immersion in domain-specific process knowledge is required for the challenging up-stream activities of formulating a coherent ML solution.

Chapter 4

Lime Kiln Monitoring with Infrared Thermal Cameras



"An infrared thermal camera monitors a unit operation in an industrial process." Text to generate images with DALL·E 2 [1].[1]

[1]Note that these are computer generated images used only for abstract conceptual visualization.

This case study explores data-driven monitoring of rotary lime sludge kilns in kraft pulping processes. The lime kiln is the single largest source of fuel consumption in pulp manufacturing [94]. One of the key motivations for this work is to address ring formation which is notorious in literature and industry for being the most troublesome problem facing lime kiln operation [126]. Visualization and data-driven analysis of shell temperature data from thermal cameras is a key focus of this research.

In the context of learning from industrial process data, the pyrometallurgy case study provides an emphasis on robust empirical comparison of traditional and contemporary process analytics techniques for predicting arc loss. The kraft pulping case study, presented in this chapter, provides a complementary perspective that emphasizes the value of process knowledge to improve the depth of interpretation while learning from industrial process data. Together, these case studies provide insights that are discussed in Chapter 5.

The material presented in this chapter is largely based on the following publications and conference presentations:

- PACWEST 2021 Technical Conference. Presentation on "Detection and Diagnosis of Ring Formation in Rotary Lime Kilns - Part I Developing a Ring Formation Indicator." Awarded Best Student Paper Award [104].

- L. Rippon, B. Hirtz, C. Sheehan, T. Reinheimer, P. Loewen, and B. Gopaluni. Visualization of multiscale ring formation in a rotary kiln. *Nordic Pulp & Paper Research Journal*, 36(4):549–558, 2021

- L. D. Rippon, B. Hirtz, C. Sheehan, T. Reinheimer, C. van der Merwe, P. Loewen, and B. Gopaluni. Detection and diagnosis of ring formation in rotary lime kilns. *Canadian Chemical Engineering Conference Proceedings*, pages 23–29, 2021

- L. D. Rippon, B. Hirtz, C. Sheehan, T. Reinheimer, C. van der Merwe, P. Loewen, and B. Gopaluni. Rotary kiln monitoring with shell temperature visualization and process analytics. In *2022 TAPPI PEERS and IBBC Conference Proceedings*. TAPPI Press, 2022

4.1 Introduction to Rotary Lime Sludge Kilns

Rotary kilns are enormous cylindrical vessels that serve as key unit operations in cement production, pyrometallurgy and kraft pulping. Industrial rotary kilns consist of a steel shell, lined with refractory bricks, rotated by a drive gear, and supported on tyres and rollers. Since these are very expensive vessels that operate at high temperatures (often exceeding 1000°C), preventative maintenance is important for rotary kiln management and safety. Moreover, the high energy intensity of rotary kilns and their fundamental role in many industrial processes provides a strong motivation to keep them operating as efficiently as possible [42].

The process monitoring techniques presented in this work are relevant to rotary kilns that collect kiln shell temperature (KST) measurements along the length of the kiln. However, investigating faults such as ring formation requires specific domain knowledge so this work focuses on a particular application, i.e., rotary lime sludge kilns in the recovery circuit of kraft pulp mills. As Figure 4.1 illustrates, rotary lime kilns (outlined in red) are part of the larger integrated forest products industry. Pulp mills are a key component of the integrated forest products industry as the profitability of sawmills depends on revenue generated through selling chips to pulp mills [15].

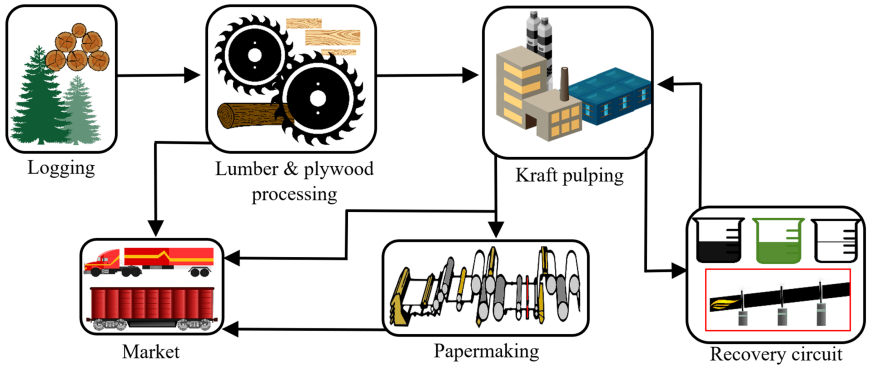


Figure 4.1: Contextualizing the lime kiln (outlined in red) within the integrated forest products industry

The kraft pulping process relies upon a strong alkaline solution known as white liquor to liberate cellulose fiber from lignin and hemicellulose during impregnation

and cooking of wood chips. An important feature of the kraft pulping process is the recovery circuit as it recovers expensive chemicals for pulp cooking and it prevents the escape of these chemicals to the environment [86]. Central to chemical recovery is the recausticizing area which produces white liquor by reacting green liquor with slaked lime. Lime mud (mainly $CaCO_3$) is a by-product of the causticizing reaction. The lime kiln is essential for regenerating burnt lime (approximately 90 wt% CaO) from this lime mud [8, 119].

Lime kilns are used in kraft pulping to regenerate calcium oxide (i.e., lime) from calcium carbonate (i.e., lime mud) according to the following endothermic calcination reaction [126]:

$$CaCO_3(s) \longrightarrow CaO(s) + CO_2(g). \tag{4.1}$$

As Figure 4.2 illustrates, wet lime mud ($CaCO_3$) is fed to the top of the inclined rotary kiln where it begins to dry into a powder before agglomerating into nodules in the preheating zone. The preheating zone facilitates an increase in temperature of the kiln solids from approximately 80°C to about 870°C at which point the calcination reaction begins [39]. Approximately 3 MJ of energy is required to produce 1 kg of pure CaO at 900°C. Rotary lime kilns can be over 4 m in inner diameter while exceeding 100 m in length [27]. The lime moves through the kiln in about 1.5 to 4 hours depending on the speed of rotation (typically between 0.5 and 2 rpm) and the slope of the kiln (between 1.5° and 3°) [4].

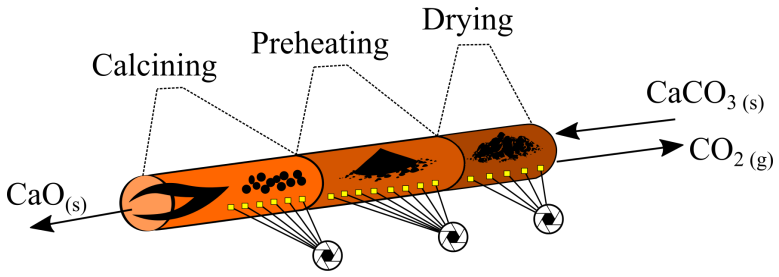


Figure 4.2: A simplified illustration of a rotary lime kiln with three thermal cameras. Lime mud is fed to the kiln where it is dried into a powder and then agglomerated into nodules which are calcined into lime product.

Another detail illustrated in Figure 4.2 is the use of thermal cameras to provide real-time KST measurements along the kiln. This work studies thermal camera data to develop kiln monitoring solutions that improve operating outcomes. This chapter discusses thermal camera applications on lime kilns and it presents two important contributions of this research, i.e., i) a novel shell temperature visualization method that provides enhanced operating insights [103], and ii) discovery of a novel phenomena, known as rotational aliasing that is a key consideration for using KST data [109]. This work empowers operators and engineers to improve the production efficiency of rotary kilns by enhancing process monitoring with thermal cameras. Although this work focuses on a lime kiln in a kraft pulping process, many of the tools and insights developed while investigating ring formation are expected to extend to other applications of rotary kilns as well.

4.2 Thermal Cameras and Data Visualization

Recently, thermal imaging tools have become significantly more affordable while offering better performance and functionality. Consequently, kilns have become increasingly equipped with infrared (IR) thermal cameras to address maintenance and operations challenges [45]. This has led to increased interest in these cameras for modeling and monitoring of rotary kilns [68, 146]. As thermal camera technology is increasingly deployed to kilns, an abundance of historical KST data is generated and the potential for data-driven optimization of kiln operating policies becomes increasingly realizable. Extracting robust insights from this deluge of historical data requires active efforts towards data storage, processing, and visualization.

A variety of devices can be used to collect KST measurements including handheld pyrometers, one-dimensional line scanners, and two-dimensional thermal cameras. The one-dimensional line scanner has a rotating head calibrated to the rotation speed of the kiln. This research studies data from a stationary two-dimensional thermal camera. These cameras can be configured to measure shell temperatures in various ways including measurement areas and profile lines, such as those shown in Figure 4.3.

Although many operations measure shell temperatures in real-time, there is often

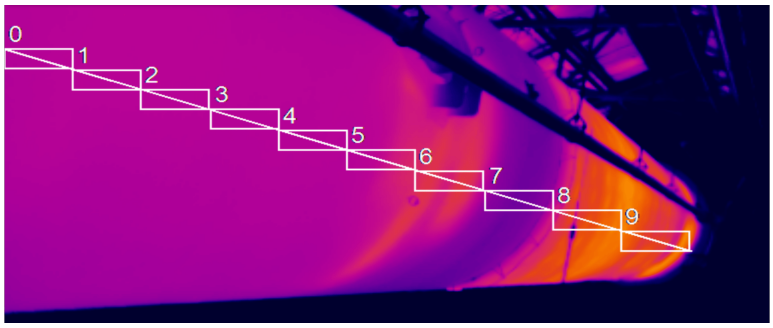


Figure 4.3: A stationary two-dimensional IR thermal camera measures shell temperatures along a rotary lime kiln [103]. The thermal camera can collect shell temperatures along the diagonal profile line or average the temperatures within the rectangular measurement areas.

significant value in the KST data that goes unappreciated and in unfortunate circumstances, becomes permanently discarded. As more kilns are equipped with thermal cameras and more industries are improving their digital capabilities, there is a valuable opportunity to develop novel data-driven solutions that tap into historically siloed sources of operating data.

4.2.1 Applications of thermal cameras on lime kilns

A variety of approaches have been introduced for deriving value from thermal camera data. Broad applications of real-time KST measurements include process monitoring, FDD, AEM, kiln modeling, and APC. These applications are studied in this work and specific details are provided in what follows. One specific application of thermal cameras is to detect refractory wear and brick loss by monitoring shell temperatures for hot-spots.

The cylindrical steel shell of the kiln is protected by a refractory which is composed of magnesia and alumina bricks with thermal and chemical resistance properties. Kiln refractories experience wear and brick loss during operation from thermo-mechanical influences, overheating, and salt infiltration. Refractory wear can be identified as hot-spots by monitoring KST data. If these hot-spots are undetected, they can cause damage to the shell which can result in catastrophic failure. Thermal cameras enable early onset detection of hot-spots so the mill can minimize damage

to the shell and avoid costly unplanned outages. Using thermal cameras to monitor refractory wear can potentially save millions of dollars in equipment damage and lost production.

In addition to identifying hot-spots and preventing damage to the kiln refractory, thermal camera data can be used to calculate shell heat loss. Radiative heat losses along the kiln are proportional to the fourth power of the shell temperatures. Shell heat losses are an important component of the kiln energy balance with a significant influence on operating efficiency and refractory design [40].

Thermal cameras can also be used to provide insight into aspects of the kiln operation such as fouling or ring formation [107]. One type of fouling, illustrated on the left side of Figure 4.4, is the formation of soda balls in the chain section of the kiln. Lime mud is washed to minimize carryover of chemicals such as NaOH which can

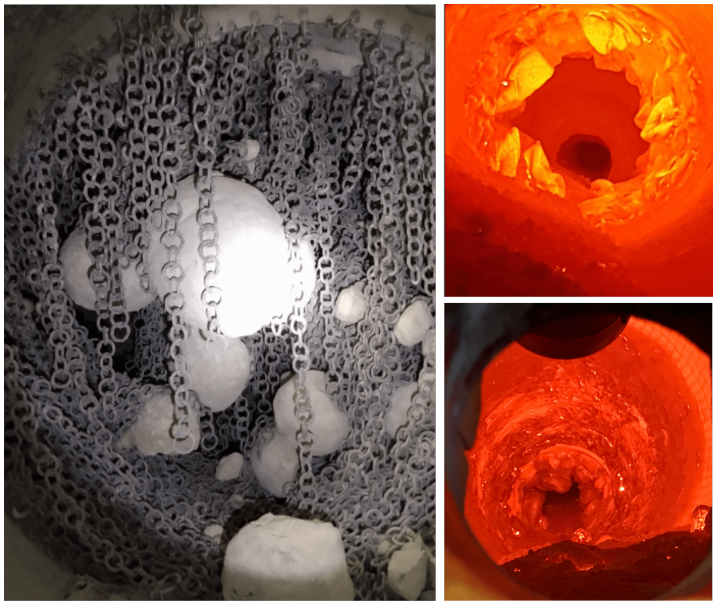


Figure 4.4: Thermal cameras can be used to monitor for kiln fouling and ring formation. **Left:** fouling in the chain section of the kiln from soda balls. **Right:** two distinct cases of ring formation [107].

volatilize in the firing end of the kiln, get carried upstream with the flue gas, condense in the chain section, and plug the kiln with soda balls. A more troublesome

type of fouling is the formation of rings on the kiln refractory, as the right side of Figure 4.4 shows. Ring formation is discussed further in Section 4.3.

Thermal cameras can also be used to evaluate shell temperature profiles which can provide important insights into the extent of the calcination reaction. High quality KST data can be used to improve the accuracy of residual calcium carbonate predictions from soft sensor models [45]. Accurately predicting the extent of the calcination reaction can enable supervisory residual carbonate control (RCC) that can minimize specific energy while producing high quality lime for slaking. Residual carbonate prediction is discussed further in Section 4.4.

In addition to measuring shell temperatures, thermal cameras can be mounted on the firing end of the kiln to monitor the shape of the burner flame, observe the size of the lime product nodules, and infer the amount of dust circulating in the kiln. However, the focus of this work is on leveraging thermal cameras that continuously measure the shell temperature profile along the length of the rotary kiln. Although process analytics (e.g., residual carbonate prediction) can help add value with thermal cameras, effective visualization is fundamental as it enables domain experts to gain operating insights from large quantities of KST data.

4.2.2 Kiln shell temperature visualization

To facilitate extraction and communication of process insights, engineers and operators must be able to interact with large quantities of thermal camera data in a user-friendly and visually intuitive manner. Bootstrapping the abilities of SMEs with enhanced data visualization is often overlooked in favor of sophisticated ML methods that attempt to conveniently bypass domain expertise. In practice, data-driven methods that neglect domain expertise fail to deliver sustained value, whereas enhancing the insights and capabilities of SMEs is significantly more effective at positively influencing operating outcomes.

One contribution of this dissertation is the introduction of a novel, user-friendly approach to visualizing large quantities of KST profiles at varying timescales. Although this visualization strategy was initially developed to support the study of ring formation, the positive reception from industry professionals demonstrated

the broader importance of this contribution as a novel tool for SMEs to troubleshoot and monitor kiln operation. The proposed visualization method has drawn interest from various mills and has been integrated into a commercial industrial software product [14].

Existing data visualization techniques

Various techniques have been used in industry and literature for visualizing shell temperatures to track ring formation. The synthetic data in Figure 4.5 illustrates a simple example of a common KST visualization technique whereby a handheld pyrometer is used to manually collect KST measurements at different locations along the kiln. Infrequent manual measurements are combined and used to construct a

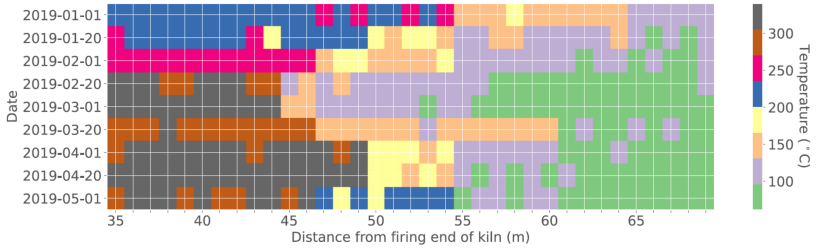


Figure 4.5: Synthetic data illustrating a low-resolution map of KST profiles.

low-resolution map of KST profiles over time with a qualitative color map [59]. The proposed approach improves this situation by enhancing the resolution, intuitiveness, and interactivity of the visualization while also collecting and processing the data using existing thermal cameras instead of manual pyrometer readings.

The continuous two-dimensional thermal camera data can be averaged over a day and stitched together to generate a KST profile as shown in Figure 4.6. The comparison of KST profiles in Figure 4.6 demonstrates a straightforward approach to visualizing ring formation. The 2019-01-01 data is from a clean kiln after a maintenance shutdown and it is compared to measurements after twenty days of operation to visualize potential ring growth between 30-45 m from the firing end of the kiln.

Alternatively, a line scanner can be calibrated with the rotation speed of the kiln to visualize the entire outer shell of the kiln over a single rotation period, as illustrated

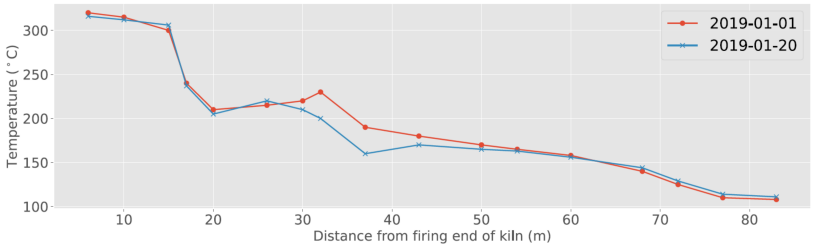


Figure 4.6: Thermal camera measurements are averaged over a day and stitched together to generate a KST profile. Measurements from a clean kiln on 2019-01-01 are compared with measurements from 2019-01-20 to emphasize potential ring formation.

by Figure 4.7. The visualization technique illustrated by the synthetic data in Figure 4.7 provides insights along the tangential dimension (or outer circumference) of the shell on the *y*-axis in addition to the axial KST variations along the *x*-axis. One drawback of this technique is the lack of a temporal dimension to provide in-

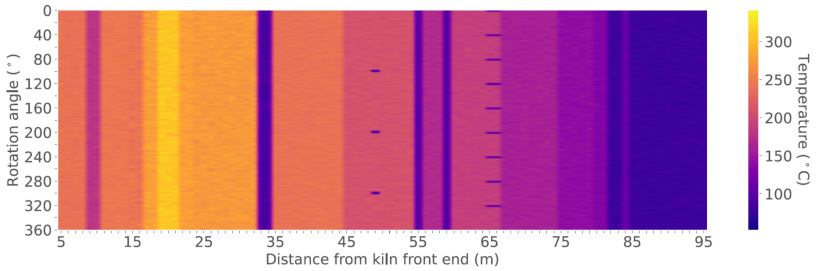


Figure 4.7: Synthetic illustration of a shell scan measured over a single rotation period. These visualizations are produced by one-dimensional line scanners that use a rotating sensor calibrated to the kiln rotation speed.

sights into the evolution of the KST profile over time (e.g., during ring formation). Also, it may become infeasible (or at least undesirable) to process and store high dimensional images for each rotation. Although shell scans from different periods can be compared to investigate changes, in practice this amounts to a cumbersome side-by-side comparison of separate images that can be overwhelming when large quantities of data need to be analyzed.

The proposed visualization strategy overcomes limitations of these existing meth-

ods by providing an intuitive visualization of large quantities of data that clearly illustrates the evolution of KSTs across any user-specified timescale. Moreover, the proposed technique is complementary to visualizations with two spatial dimensions (e.g., Figure 4.7 or raw camera images). The proposed technique can be used to identify periods of interest from which to investigate further with raw images.

Spatiotemporal heatmap for scalable shell temperature visualization

The method presented here includes the conversion of raw KST data from a thermal camera to a visually intuitive spatiotemporal heatmap. This is followed by the development of a user-friendly interactive approach to controlling the timescale at which kiln temperature dynamics are observable.

Consider a matrix of KST measurements, $\boldsymbol{T}(\boldsymbol{x},\boldsymbol{t})$, where the rows of $\boldsymbol{T}$ are indexed by $\boldsymbol{x} = [x_1, x_2, \ldots, x_n]$ and the columns of $\boldsymbol{T}$ are indexed by $\boldsymbol{t} = [t_1, t_2, \ldots, t_\tau]$. Each x_i represents the distance from the firing end of the kiln that the thermal camera measurement is taken whereas each t_i represents a periodic sequence of discrete sampling times with a total of τ samples. Each row, i, of $\boldsymbol{T}$ is a univariate time series, $T(x_i,\boldsymbol{t})$, composed of τ total KST measurements at position x_i from the firing end of the kiln. Each column, j, of $\boldsymbol{T}$ is a snapshot of the entire KST profile at time t_j, denoted $T(\boldsymbol{x},t_j)$.

The KST measurement positions, $\boldsymbol{x}$, are determined by the placement and configuration of the thermal camera(s) which may result in positions that are not equispaced. If these non-equispaced positions are not addressed they can create visual obscurities which can cause misleading conclusions about the spatial dynamics in the kiln. A snapshot of a non-equispaced KST profile is presented in Figure 4.8. Each temperature in Figure 4.8 is assigned a color, as shown in the color bar on the left. Measurements are colored accordingly, effectively embedding the y-axis as uniform, sequential colors, as a precursor to further refinements below.

To address the non-equispaced samples an upsampled KST profile is created, $\tilde{\boldsymbol{T}}(\boldsymbol{x},\boldsymbol{t})$, by performing piecewise linear interpolation on positions between those provided in the original row index, i.e., $x_i < x < x_{i+1}$. Since the interpolation is only along the spatial dimension, $\boldsymbol{t}$ is dropped for notational simplicity. The simplest piece-

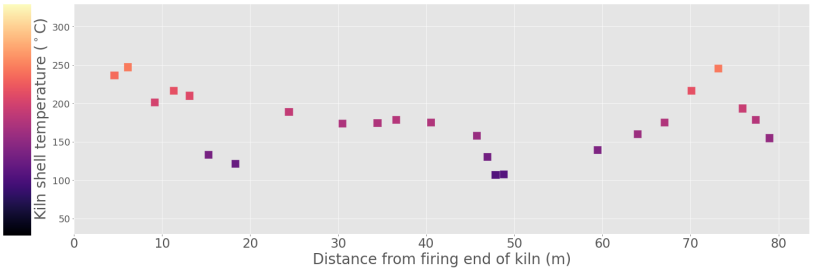


Figure 4.8: A KST profile with non-equispaced measurement positions. Discrete measurement areas summarize thermal camera pixels into an average temperature at fixed positions and intervals.

wise linear interpolation function defined on $[x_1, x_n]$ that reproduces the measured temperatures is

$$\tilde{T}(x) = T(x_i) + \frac{x - x_i}{x_{i+1} - x_i}[T(x_{i+1}) - T(x_i)] \quad \text{for} \quad x_i < x < x_{i+1}. \tag{4.2}$$

Repeating this interpolation for all positions defined on $[x_1, x_n]$ yields the upsampled KST profile which is presented for the same sample time in Figure 4.9. Once

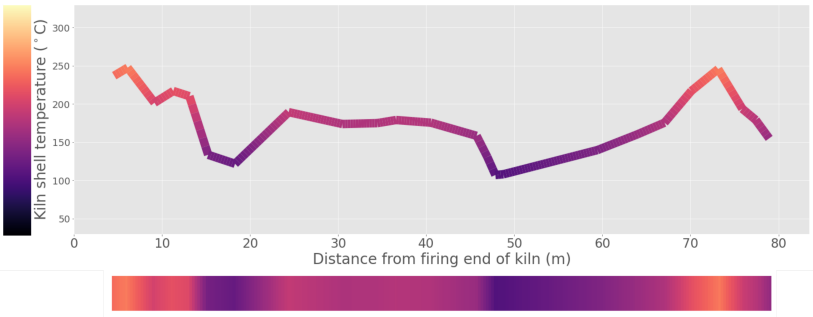


Figure 4.9: Piecewise linear interpolation of the KST profile. The bottom color bar embeds the y-axis temperatures with a uniform, sequential colormap enabling a clear and intuitive single-axis visualization.

again, the y-axis values map directly to the vertical color bar legend on the left-hand side. The horizontal image at the bottom of Figure 4.9 represents the same KST profile but with the y-axis embedded in the colors according to the legend. Rotating this image clockwise by 90° provides one column from the upsampled KST

matrix $\tilde{\boldsymbol{T}}(\boldsymbol{x},\boldsymbol{t})$. Repeating this process for all samples yields an image of $\tilde{\boldsymbol{T}}(\boldsymbol{x},\boldsymbol{t})$ that enables a visually intuitive representation of large quantities of KST data.

For consistency, it is recommended to select a colormap that resembles the raw thermal camera images. Once measurements with invalid magnitudes are addressed, the maximum and minimum values of $\tilde{\boldsymbol{T}}$ are used as limits for the colormap. Intermediate values are mapped in a uniform fashion with equal increases in temperature yielding equal increases along the colormap. The resulting visualization is a spatiotemporal heatmap of the kiln over τ sampling periods. Figure 4.10 shows a spatiotemporal heatmap with over 5 years of hourly averaged KST profiles.

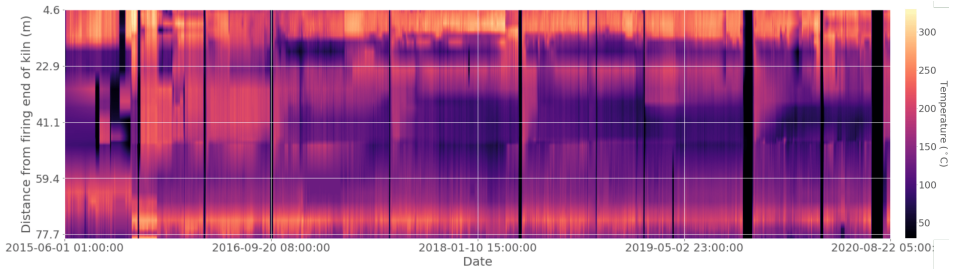


Figure 4.10: Spatiotemporal heatmap of KST profiles spanning over 5 years of operation. This is the first published visualization that offers a clear and intuitive view of shell temperature evolution over years of rotary kiln operation.

If the time scale, τ, is too large then high frequency variations are compressed into a small number of pixels rendering them incomprehensible. To visualize the KST profile at varying time scales an interactive heatmap is developed with user-specified parameters for the start date (s) and the window size (w). Ultimately, an upsampled slice of the raw KST matrix, i.e., $\tilde{\boldsymbol{T}}(\boldsymbol{x},t_{s:s+w})$, is used to analyze the KST profile of the kiln. The following material demonstrates the utility of the proposed visualization strategy with an industrial case study.

Industrial case study

Consider an industrial rotary lime kiln in a kraft pulping process that is approximately 85 m long and 3 m in diameter. Like many kilns, this kiln is challenged by

the formation of rings. For improved control and management of abnormal situations (such as rings) this kiln has been equipped with three thermal cameras that measure the shell temperature profile at non-equispaced locations along the kiln.

The imaging resolution for each thermal camera is 480 $\times$ 360 (width $\times$ height) temperature measurements taken 30 times per second. To satisfy historian storage constraints the axial width of each measurement area is set to approximately 3 meters (although this can be adjusted to be larger or smaller) and the data is reported to the historian every minute. The thermal cameras are capable of measuring temperatures between -40°C to 1500°C with an accuracy of $\pm\ 2^{\circ}$C.

Each camera logs ten measurements in the historian, but two measurements from each camera are omitted from the heatmap because they are related to the external bracing around the kiln. A total of 24 KST measurements are sampled along the length of the kiln creating a shell temperature profile with measurements spanning between 4 to 79 meters from the firing end of the kiln. To analyze the formation of rings at varying time scales a dataset of KST measurements and other relevant PVs is acquired using hourly averages from over 5 years of historical operation. The matrix of raw KST profiles ($\boldsymbol{T}$) are upsampled and the resulting matrix ($\tilde{\boldsymbol{T}}$) is used to generate the spatiotemporal heatmap in Figure 4.10.

The *y*-axis of the heatmap shows the distance from the firing end of the kiln in meters while the *x*-axis provides the date from the sample timestamp. A legend on the right side of Figure 4.10 shows the relationship between colors and shell temperatures. An intuitive correlation of temperature with brightness provides immediate operating insights, e.g., the vertical black slices are from prolonged maintenance shutdowns. The period to the right of 2018-01-10 in Figure 4.10 shows a shutdown followed by a period of presumed fouling at approximately 20-40 m from the firing end of the kiln. The fouling is suspected because the shell becomes cooler (i.e., the heatmap becomes darker) in that section while remaining relatively stable at both the firing end and feed end of the kiln. To better investigate ring formation, the start date and window size can be specified to zoom in on this event as shown by the heatmap in Figure 4.11.

Reducing the window size reveals insights into this period of ring growth. Fig-

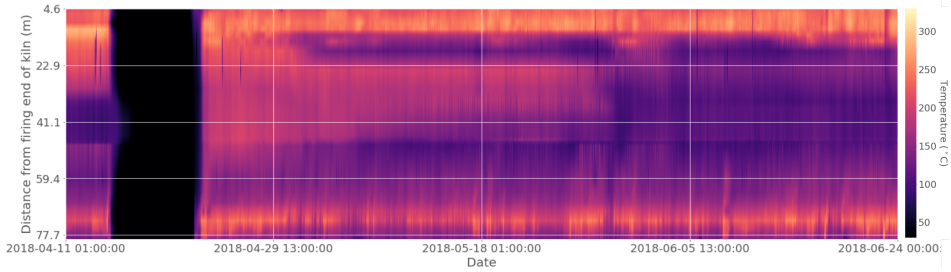


Figure 4.11: Focusing on a period of suspected fouling. Potential ring formation is observed after 2018-04-29 at approximately the 18 m and 45 m positions. Between 2018-05-18 and 2018-06-05 the fouling appears to bridge these separate formations and extend into the 23-40 m range.

ure 4.11 shows that a ring appears to form after 2018-04-29 at approximately 18 m from the firing end of the kiln. Also, a much longer ring appears to form between approximately 23-40 m from the firing end of the kiln. The presumed formation of this longer mid-kiln ring happens between 2018-05-18 and 2018-06-05. Given these specific periods of interest, further investigation can be undertaken, e.g., zooming in further with the heatmap or directly inspecting the raw thermal camera images. In this case the latter choice is selected to demonstrate how the proposed visualization strategy complements the raw thermal camera images.

Figure 4.12 compares two images taken from the thermal camera on the firing end of the kiln. The image on the left side of Figure 4.12 is taken from 2018-04-29 while the image on the right side of Figure 4.12 is taken from 2018-05-06. Note that the right-side image exhibits significantly darker shell colors in measurement areas E, F, and G, which corresponds to the shell temperature decrease observed in Figure 4.11.

Figure 4.13 investigates the longer, mid-kiln temperature drop by comparing raw images taken from the thermal camera that monitors the middle section of the kiln. The image on the left side of Figure 4.13 is taken from 2018-05-23 whereas the image on the right side is taken from 2018-05-30. During this week of operation, measurement areas C, D, E, and H become significantly darker which corresponds to the drop in temperature observed in Figure 4.11 both in terms of location and

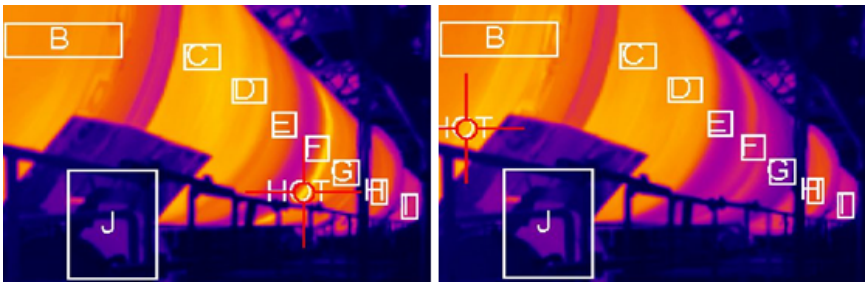


Figure 4.12: Investigating thermal camera images to confirm suspected ring formation approximately 18 m from the firing end of the kiln just after 2018-04-29. **Left:** an image from the firing end camera taken on 2018-04-29. **Right:** an image taken on 2018-05-06 with cooler shell temperatures in measurement areas E, F, and G.

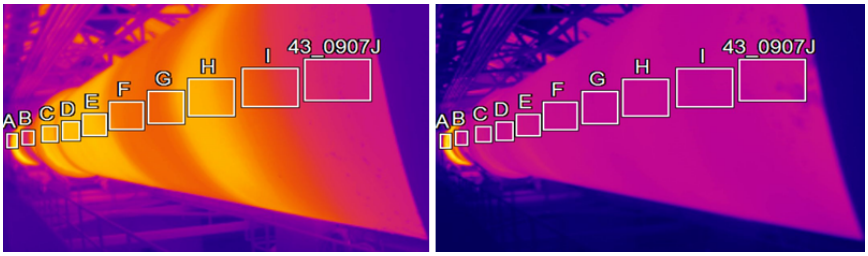


Figure 4.13: Observing mid-kiln camera images to confirm ring formation in the 23-40 m range between 2018-05-18 and 2018-06-05. **Left:** image taken on 2018-05-23. **Right:** image from 2018-05-30 that exhibits cooling across all measurement areas except area A.

time period. The consistent temperature drop shown on the right side of Figure 4.13 is not necessarily indicative of ring formation. Further investigation can be conducted to determine potential explanations for the shell temperature changes from operating data.

To supplement the heatmap, another feature is included that allows users to select from a list of PVs to plot. Axes, units, and descriptions are updated automatically. This feature is demonstrated in Figure 4.14 to investigate the cause of the long and consistent decrease in shell temperatures found in Figure 4.13. The solids percentage of the calcium carbonate lime mud fed to the kiln is shown in the bottom plot of Figure 4.14. A sustained drop in mud solids between 2018-05-28 and 2018-

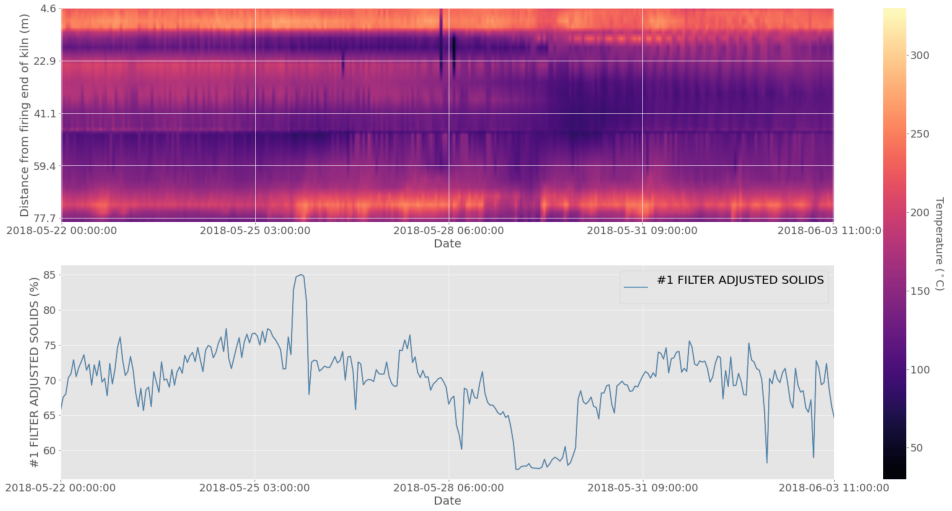


Figure 4.14: Further reducing the heatmap window and plotting relevant PVs for operating context. **Top:** the spatiotemporal heatmap of KST profiles. **Bottom:** the solids percentage of the lime mud fed to the kiln.

05-31 suggests the observations from the right side of Figure 4.13 are due to poor performance of the mud filters resulting in lime mud that is too wet. Mill records confirm mud filter problems during this period. Moreover, mill personnel have observed a seasonal pattern where spring runoff results in an increase in mill water turbidity and perhaps a reduction in mud filter performance.

In this case study, a ring formation event is investigated using the spatiotemporal heatmap method for visualizing shell temperatures. A ring is identified at roughly 18 m from the firing end of the kiln on 2018-04-29. Mid-kiln cooling is observed between 23 m and 40 m from the firing end of the kiln after an extended decrease of mud solids on 2018-05-29. Mud filter performance is well-known as a critical factor for ensuring proper functionality and efficiency of lime kilns [90]. Although the feed end temperatures recover from this moisture upset, the mid-kiln temperatures at approximately 38 m from the firing end of the kiln do not recover. Ultimately, a ring forms in this area, and it needs to be manually removed during a shutdown.

The convenient and intuitive nature of this visual tool enables navigation and comprehension of large quantities of high dimensional historical data. In addition to

identifying periods of fouling the heatmap can help evaluate data quality and the effect of data pre-processing activities. Individual shutdowns can be identified, and start-up procedures can be compared. The state of the process can be succinctly compared across years of operation. Plotting suspected variables alongside the heatmap helps to identify poor pre-coat filter performance as a cause of ring formation. It is conceivable that dust in the flue gas contacts the excess moisture and forms deposits along the kiln refractory. The following section provides additional information on ring formation. The heatmap visualization code is freely available along with a brief instructive tutorial using synthetic data [102].

4.3 Monitoring Ring Formation

The previous section introduced a novel shell temperature visualization method and demonstrated its effectiveness for monitoring and investigating ring formation events. In this section process analytics and ML are studied for monitoring ring formation with thermal cameras and traditional instrumentation. First, background information on ring formation is provided.

4.3.1 Ring formation in lime sludge kilns

Reaction materials flow through the kiln because of the slope and rotation speed of the kiln. Apart from a thin coating that is applied to the refractory during start-ups, there should be no accumulation of material in the kiln. The formation of rings, such as those shown in Figure 4.15, is a significant outstanding problem in the operation of rotary kilns. An inquiry on Swedish kraft pulp mills showed that approximately 70% of mills suffered from ring formation and many did not know why [77]. Rings occur when there is adhesion and accumulation of lime mud or product lime particles to the refractory wall on the inside of the kiln. This results in the formation of rings and annular cylinders that can restrict the flow of gas and solids in the kiln. Formation of rings is the most troublesome problem for lime kiln operation [126].

Rings can form at significantly different rates, with noticeable deposits accumulating quickly over a matter of days, or gradually over a matter of months. This creates

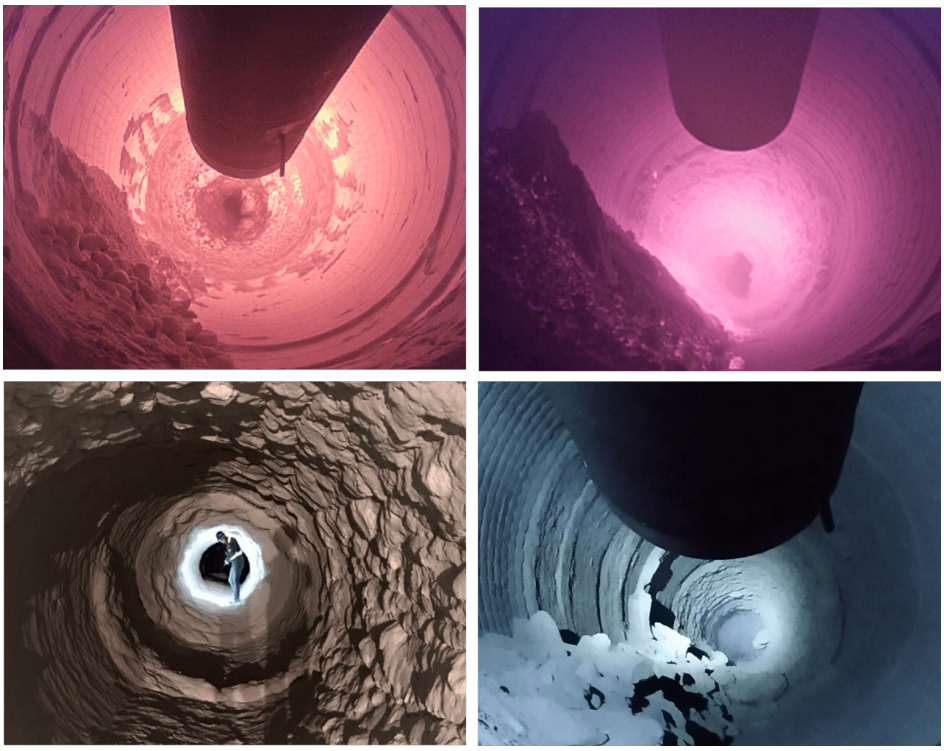


Figure 4.15: Ring formation in rotary lime sludge kilns. **Top:** severe ringing restricting the orifice of the kiln in the mid-zone and firing end. **Bottom left:** a mill worker studies ring formation inside a kiln. **Bottom right:** stress fracturing of ring into debris with thermal cycling.

challenges for detecting and visualizing ring growth. The ring orifice can reduce lime production costing over $50,000 per day in purchased lime rock. Moreover, if the ring goes unnoticed it can result in overheating and damage to the refractory lining which may require repairs and lost production in excess of $3 million per event. Although the exact mechanisms of ring formation are not completely understood, distinct types of rings have been observed and causal mechanisms have been proposed [39].

Rings can form at different locations along the length of the rotary lime kiln and for various reasons. For example, rings near the firing end are often attributed to

sulphation reactions, i.e.,

$$\mathrm{CaO(s)} + \mathrm{SO_2(g)} + \frac{1}{2}\mathrm{O_2(g)} \longrightarrow \mathrm{CaSO_4(s)} \tag{4.3}$$

where the $\mathrm{CaSO_4(s)}$ product hardens the ring, making it more resistant to abrasion. The presence of $\mathrm{SO_2(g)}$ is due to impurities in either the fuel or the lime mud [128].

Mid-kiln rings are often the most troublesome and they form as a result of carbonation which is the reverse of calcination in reaction (4.1) [127]. Carbonation, which can significantly harden ring deposits, has been attributed to unstable operation and fluctuating temperatures [31]. Moreover, an unfortunate positive feedback loop has been studied whereby the growth of rings contributes to temperature fluctuations by affecting the shape and stability of the flame [87].

Softer mud rings that develop near the feed end of the kiln are believed to be caused by high mud moisture and low temperatures. High sodium content (e.g., $\mathrm{Na_2CO_3(aq)}$) in the lime mud as a result of poor washing and low solids content can result in rapid deposition of sticky lime to the kiln refractory lining [98, 126].

Mills attempt to mitigate ring formation by using industrial 8-gauge shotguns, thermal cycling to cause stress fractures, adding water to the kiln (i.e., slaking) to dissolve and soften the ring, and blasting with high pressure carbon dioxide (i.e., cardox) [39]. For complete removal it is often necessary to shut down the kiln and wait for it to cool enough for someone to enter the kiln and use a pneumatic jack hammer to physically remove the ring. Early detection of rings through effective kiln monitoring is paramount for minimizing their impact on operations and identifying process conditions that lead to ring formation.

First principles mathematical modeling coupled with laboratory experiments provides valuable insights into the complicated mass and energy balance dynamics in rotary lime kilns [5, 32, 36, 37, 83]. Heat transfer modeling has been combined with kiln shell scanning to estimate the thickness of the refractory coating [88]. However, given the complexity of factors surrounding ring growth as well as complications such as refractory wear, implementation of these models on existing operations can be challenging. Instead, this work aims to use ML methods and

statistical analysis to monitor ring formation and diagnose potential causes.

4.3.2 Developing a ring formation indicator

In literature, fault detection is often reduced to binary classification, i.e., whether or not a fault has occurred. Like-wise, diagnosis is often reduced to multi-class classification, i.e., which fault occurred. These experiments are dependent on idealistic labelled datasets where there is no ambiguity regarding the ground-truth of the class labels [81]. In practice, when dealing with outstanding industrial faults and real historical process data, the situation is often far less ideal. In this case, there is no labelled dataset that says where and when faults occurred and what the exact cause was. Instead, as is discussed in what follows, most of the work involves developing techniques to reliably label the historical data.

The use of data-driven FDD strategies to mitigate the negative impacts of ring formation has received little attention in literature. The primary objective of ring detection is to provide early indications of ring formation so corrective actions can be taken by operators. A secondary objective of ring detection is to label five years of historical data such that supervised learning methods can be used for identifying high-risk operating conditions and diagnosing ring formation. To accomplish both objectives a ring formation indicator is developed and validated with KST data and raw thermal camera images.

The proposed ring detection algorithm is based on monitoring residuals (e) between estimated temperatures ($\hat{T}$) and observed temperatures (T). Consider a sample time t as an element of a sequence of sample times $t = (t_1, \ldots, t, \ldots, t_\tau)$ and a specific position x along the KST profile $x = (x_1, \ldots, x, \ldots, x_n)$ measured in terms of distance from the firing end of the kiln. The residual is computed as follows:

$$e(x,t) = \hat{T}(x,t) - T(x,t) \tag{4.4}$$

where high positive values of the residual indicate a lower-than-expected measured temperature (i.e., potential ring growth), and high negative values indicate a higher-than-expected measured temperature (i.e., potential ring decay).

Attributing shell temperature residuals to ring formation assumes the mismatch between estimated and actual temperatures is due to rings causing changes in the thermal resistance of the kiln shell wall. As Section 4.2 demonstrates, the presence of disturbances (e.g., moisture upsets) and confounding variables (e.g., refractory wear) make this a precarious assumption. Therefore, special consideration must be given to predicting temperatures and validating residuals.

Shell temperature forecasting and residual monitoring

A straightforward method to create shell temperature residuals is to use a reference KST profile after the kiln is cleaned during a shut-down. However, as Figure 4.10 demonstrates, the relevance of a post-startup reference KST profile is brief.[2] Moreover, ongoing refractory wear, changing firing rates, and process upsets can confound ring growth indications and further trivialize dated reference profiles. To address these challenges, a statistical forecasting technique is designed with process knowledge to control for slow KST changes (e.g., due to refractory wear) and expected shell temperature changes (e.g., due to firing rate changes).

Initially, a simple linear time-series model is proposed. Consider a first order autoregressive (AR) model, denoted AR(1), where at a specific position x along the kiln the shell temperature at time t is modeled as

$$T(x,t) = \phi_0 + \phi_1 T(x,t-1) + \varepsilon(t), \tag{4.5}$$

and the predicted temperature at time $t+1$ is given by

$$\hat{T}(x,t+1) = \phi_0 + \phi_1 T(x,t), \tag{4.6}$$

where ϕ_0 and ϕ_1 are unknown parameters that are estimated during training. The AR model can be seen as a starting point for linear time-series modeling which can be easily extended into more advanced linear models.

Consider a model where the endogenous response variable, $\boldsymbol{T}(t)$, is a vector of the entire KST profile at time t, i.e., $\boldsymbol{T}(t) = (T(x_1,t), T(x_2,t), \ldots, T(x_n,t))^\top$. To

[2] Many useful insights for developing a ring formation indicator are available in Figure 4.10.

model shell temperatures, a vector auto-regressive moving-average exogenous input (VARMAX) model is used with a variable AR order p, and a variable moving-average (MA) order q. This is known as the VARMAX(p,q) and it models the KST profile at time t as

$$\boldsymbol{T}(t) = \phi_0 + \sum_{i=1}^{p} \boldsymbol{\Phi}_i \boldsymbol{T}(t-i) + \sum_{i=1}^{q} \boldsymbol{\Theta}_i \boldsymbol{\varepsilon}(t-i) + \sum_{i=0}^{b-1} \boldsymbol{B}_i \boldsymbol{X}(t-i) + \boldsymbol{\varepsilon}(t), \qquad (4.7)$$

where ϕ_0 is a vector of unknown constants, $\boldsymbol{\Phi}_i$ is a matrix of AR coefficients, $\boldsymbol{\Theta}_i$ is a matrix of MA coefficients, $\boldsymbol{B}_i$ is a matrix of exogenous coefficients, and $\boldsymbol{\varepsilon}(t)$ is a multivariate extension of $\varepsilon(t)$ from equation 4.5. There is a small discrepancy in notation to be aware of, capital $\boldsymbol{X}(t)$ is a vector of exogenous variable observations whereas lowercase x denotes the position along the kiln. A total of b observations for each exogenous variable are considered including those at the current time.

Exogenous regressors can be variables like the flow of fuel to the burner or the flow of lime mud to the feed-end of the kiln. A common pitfall in process monitoring applications of ML (as discussed in Chapter 5) is that domain knowledge is disregarded and selection of exogenous regressors is reduced to a data-driven endeavour to optimize a performance metric, e.g., root mean square error (RMSE). However, in practice it is often beneficial to select exogenous regressors with careful regard for the operating behaviour they help represent. Models that represent the expected behaviour of the process can be particularly useful for fault detection.

Models of this nature are trained on the historical KST data to generate temperature forecasts with a rolling origin as demonstrated by Figure 4.16. Taking the difference of the forecasted and measured temperatures results in a series of residuals. The large drops in measured temperature in Figure 4.16 demonstrate a problem with the raw camera data, i.e., intermittent obstructions of the camera as shown in Figure 4.17. Figure 4.17 is an important reminder that it is critical to understand the quality of process data for ML applications (e.g., APC, FDD, etc.). Maintenance of instrumentation to produce high quality data in large-scale industrial processes is a significant challenge. Practical constraints and human error can result in sub-optimal installation, configuration, and maintenance of process instrumentation.

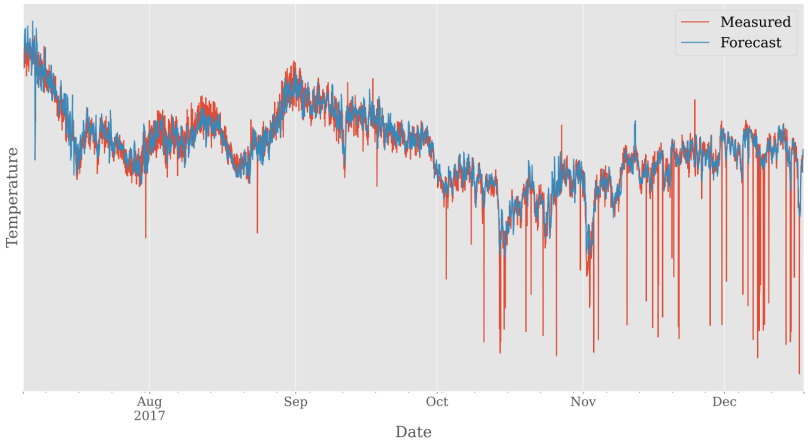


Figure 4.16: Statistical forecasting of shell temperatures from a thermal camera at a specific axial position along the kiln [107].

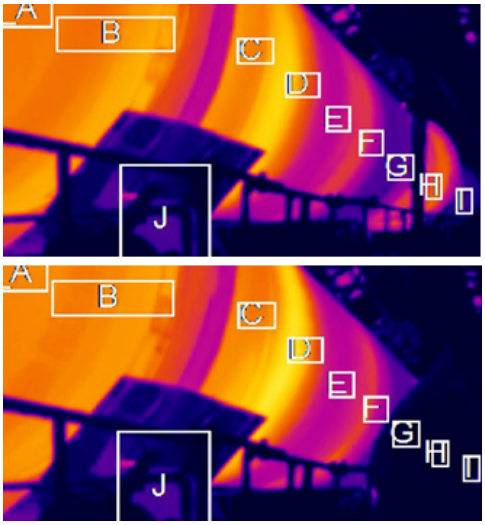


Figure 4.17: Obstruction of measurement areas G, H, and I in the bottom image can corrupt the thermal camera data.

To obtain more robust indications of ring formation, a simple thresholding procedure is used to isolate shell temperature residuals that are considered significant. Each series of residuals is analyzed as a distribution and the thresholds are based on the first quartile, third quartile, and the inter-quartile range (IQR) of the distri-

bution. Growth residuals are greater than one IQR above the third quartile, and less than three IQRs above the third quartile. Similarly, residuals are considered indications of ring decay if they are greater than three IQR below the first quartile and less than one IQR below the first quartile.

Once the significant growth and decay residuals are sampled, daily counts for each growth and decay residual are generated as shown in Figure 4.18. A final threshold is applied to the daily counts to identify days where ring formation is suspected. Days that are flagged for ring formation are validated with raw data.

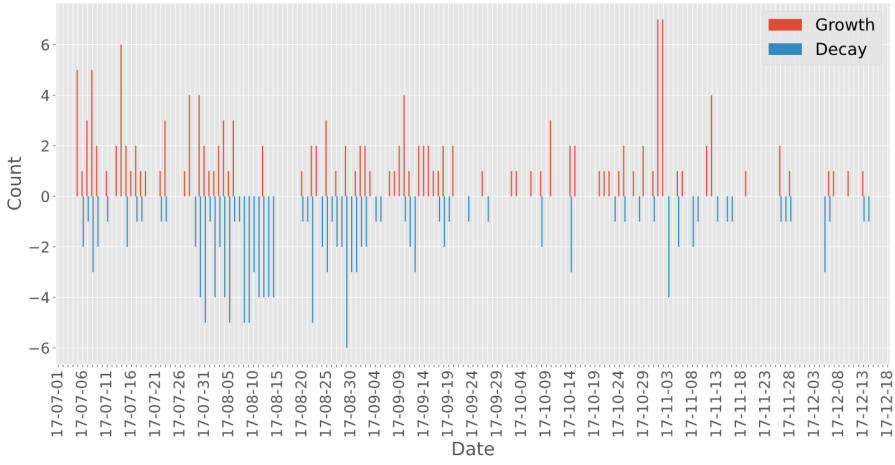


Figure 4.18: Daily counts for growth and decay indicators based on shell temperature residuals.

As mentioned before, an important prerequisite for using supervised learning to predict or diagnose ring formation is an accurately labelled dataset. The primary quantitative result of this study involves validating and improving the reliability of ring detection. The heatmap visualization and raw thermal camera images are studied to determine the accuracy and utility of the proposed ring detection method. Troubleshooting and validating the ring detection algorithm leads to the discovery of a novel phenomena known as rotational aliasing.

Validating the ring formation indicator

The analysis begins with five years worth of hourly averaged KST measurements from three thermal cameras. A snapshot of the image from the firing end camera is shown in Figure 4.17. This data contains unwanted periods from shutdowns and sensor failures which appear in Figure 4.10 as vertical streaks and periods with zero variability, respectively. Extracting periods where the entire KST profile is valid for at least 30 days results in twelve distinct periods of operation, referred to as experimental trials.

Although there are 24 measurement positions along the KST profile, for now consider just a single position, i.e., 18 m from the firing end of the kiln. A univariate formulation of equation 4.8 is used to forecast shell temperatures at the 18 m position. The IQR filtering is conducted to sample significant residuals from which daily counts are generated. Figure 4.16 and Figure 4.18 demonstrate the forecast and the daily counts at 18 m for the longest experimental trial. This procedure is repeated for each of the twelve experimental trials and a threshold is applied to all 751 days. Days with growth/decay counts greater than 4 are flagged for ring growth/decay, respectively. Ultimately, 24 days are flagged for ring growth and 51 days are flagged for ring decay. Manual validation is conducted for both growth and decay using the raw thermal camera images and other available resources.

Consider the period around 17-11-03 from Figure 4.18 which shows two high growth days followed by one decay day. To validate these events we observe the raw thermal camera images (focusing on measurement area G) and compare the forecast to the measured temperature as shown in Figure 4.19. The growth indication on 17-11-01 is considered a TP given the clear growth from the day prior. For similar reasons the decay indication on 17-11-03 is also considered a TP. However, the growth indication on 17-11-02 is not as certain. The temperature trends indicate potential growth in the first half of the day, but this is not clear from the thermal camera images which are only available every four hours. Given the limitations on the available data, efforts towards validating ring detection suffer from inherent ambiguity and subjectivity. As discussed in Chapter 5, this is not a recipe for successfully applying ML to industrial data.

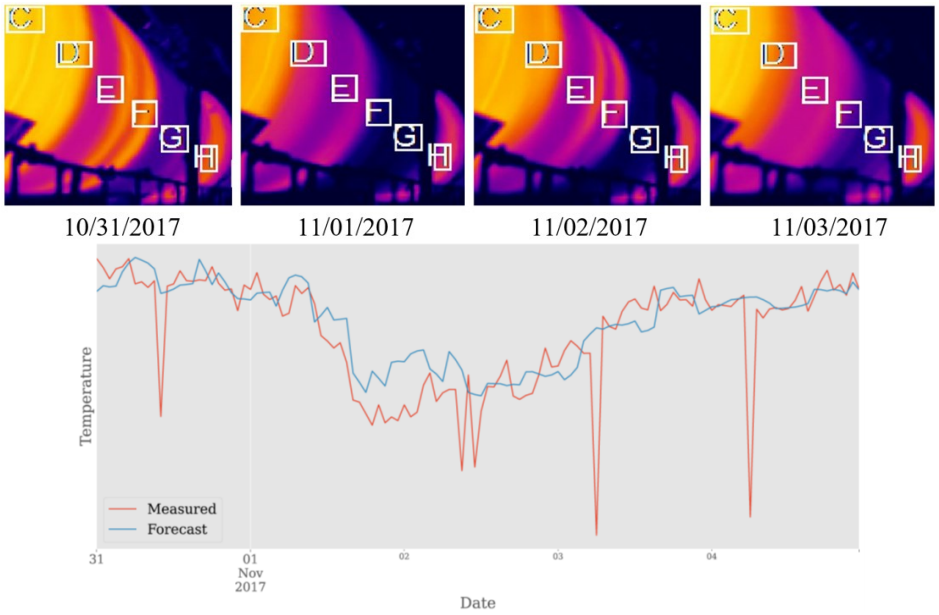


Figure 4.19: Validating ring formation indications with raw thermal camera images (top) and forecast results (bottom).

A similar manual validation procedure is performed for each of the 75 flagged days. In binary classification terms, two separate binary classifications are performed, i.e., one for ring growth and one for ring decay. The flagged days represent positive events, and the validation of these days is performed to determine the precision of each indicator, i.e., the ratio of TP to all positive events. Table 4.1 shows the results of the manual ring indicator validation. To address events with insufficient

Table 4.1: Validating growth and decay indications for precision.

Indication	Precision	Certainty
Growth	68.2%	31.8%
Decay	87.5%	42.5%

evidence a second binary label is applied to each event, i.e., whether the validation assignment is certain. The assignment of certainty is itself subjective, but it provides valuable insight into the ambiguity involved with assigning ground truth

labels to the available data.

As Table 4.1 shows, the precision for the growth indicator is 68.2% but only 31.8% of validation labels are considered certain. The results for ring decay are slightly better with a precision of 87.5% and a certainty of 42.5%. The major sources of uncertainty are insufficient raw data (e.g., thermal images are too infrequent) and rotational aliasing. Given this untenable degree of uncertainty, improvements are required to proceed with ring detection. In what follows, the discovery of rotational aliasing is described and a straightforward solution is implemented to improve the quality of KST measurements.

4.3.3 Rotational aliasing

Attempting to manually validate the ring formation indicator with historical data led to the discovery of a novel phenomena referred to here as rotational aliasing. The underlying principles causing rotational aliasing are similar to the aliasing of machine direction and cross direction variations in paper machine control [105]. This research represents the first time that rotational aliasing of shell temperature data on rotary kilns has been formally introduced [109].

Using statistical forecasting to analyze shell temperatures led to the discovery of large magnitude oscillatory residuals. Further analysis showed these residuals to be a result of the underlying KST data. The high frequency shell temperature variations observed in Figure 4.20 were initially puzzling especially given the somewhat regular periodicity and the magnitude of the shell temperature variations (exceeding $\pm 50^{\circ}$C per hour). Given that the variations appear near the firing end of the kiln, one proposed mechanism was the impingement of the burner flame on the refractory wall. Further investigation proved that these shell temperature variations are due to aliasing as a result of the measurement configuration.

Prior to discovering this high frequency variability, it was falsely assumed that each datapoint was roughly representative of the average temperature along the entire circumference of the shell at the given axial position. On the surface this appeared to be a reasonable assumption given the rotation speed of the kiln is just over one minute and the KST data is averaged hourly. The heatmap in Figure 4.20

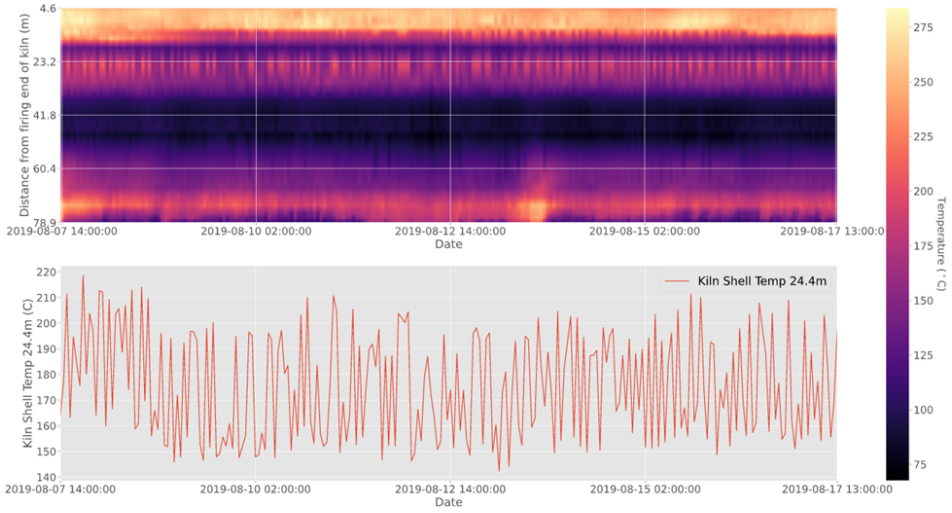


Figure 4.20: High frequency variations in shell temperatures measured 24.4 m from the firing end of the kiln. Spurious oscillations corrupt the KST profile and hinder efforts towards ring detection.

gives the false impression of a ring rapidly growing and decaying approximately 23.2 m from the firing end of the kiln. Observing the 50°C hourly oscillation in the plot below the heatmap led to uncertainty regarding the assumption that each hourly averaged datapoint accurately represented the average temperature over the entire shell circumference.

Further investigation into the data collection and processing revealed that the hourly averaged data is drawn from the process historian which averages from an internal database that stores KST data once per minute. However, the one-minute frequency data is not a result of averaging even higher frequency measurements taken from the camera. Instead, the one-minute data in the process historian is collected as an instantaneous snapshot of the kiln shell that is within the measurement area.

If the kiln is rotating at around 1.25 rpm (as this kiln is) then the frequency of a disturbance (e.g., fouling) at one spot along the inner circumference of the kiln is roughly 0.021 Hz. From the Nyquist-Shannon sampling theorem the minimum sampling rate required to perfectly reconstruct a signal with a frequency of 0.021 Hz is 0.042 Hz, i.e., one sample every twenty-four seconds [115]. In other words,

the KST data stored in the historian did not meet the Nyquist rate of 0.042 Hz and therefore this data is insufficient for reconstructing temperature variations associated with the rotation of the kiln, e.g., fouling that is not uniformly distributed along the inner circumference of the kiln.

A new KST sampling strategy is proposed such that a profile line (e.g., the diagonal line in Figure 4.3) is used to measure the shell temperatures and record them every five seconds. To manage the increased spatial and temporal resolutions the five-second samples are subsequently averaged into five-minute samples. The new KST profile line strategy is implemented for one of the three thermal IR cameras, and it consists of 380 measurements located roughly 2 m to 25 m from the firing end of the kiln. This high resolution KST data is demonstrated as a heatmap in Figure 4.21. Given that the process historian continues to collect KST data as one-minute

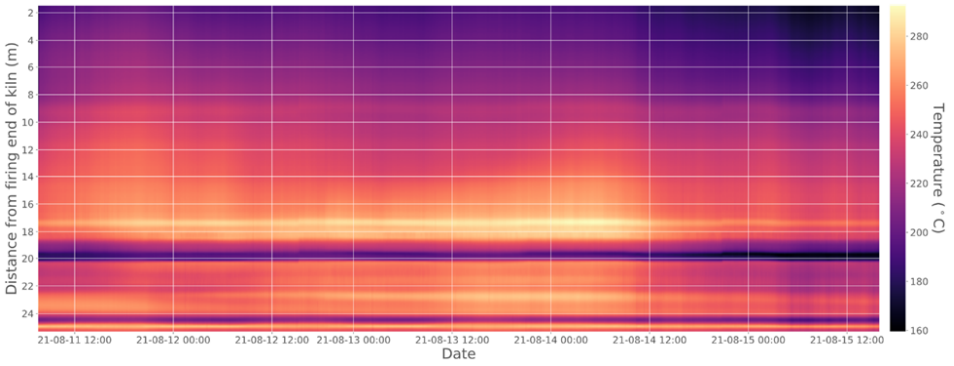


Figure 4.21: Spatiotemporal heatmap from the thermal camera on the firing end of the kiln providing improved spatial and temporal resolution.

snapshots over the rectangular measurement areas, it is sensible to compare the two sets of data to see if the high frequency noise is attenuated.

Since there are only eight KST measurement areas from the firing end camera the comparison relies on determining which of the 380 profile line measurements best correspond to the rectangular measurement areas. Fortunately, the thermal camera software provides some assistance. The profile line data is distinguished by the x-coordinate of each pixel in the camera's field of vision. Each measurement area has a center x-coordinate that is given by the thermal camera software. For example,

the center x-coordinate of the 18.3 m measurement area is 328. Therefore, profile line trends with x-coordinates between 328 ± 30 are collected and correlated with the KST data from the 18.3 m measurement area.

Figure 4.22 shows the resulting correlation coefficient for each of the profile line positions. A maximum correlation coefficient of 0.972 is obtained at x-coordinate 325. Figure 4.23 shows the comparison of the measurement area data (red) with

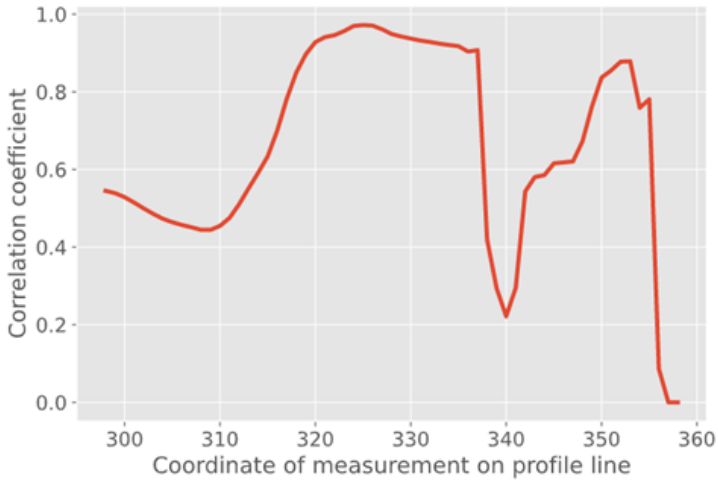


Figure 4.22: The correlation coefficient between KST data from a rectangular measurement area 18.3 m from the firing end of the kiln and KST data from various profile line positions denoted by their x-coordinate along the profile line.

the profile line data (blue) which demonstrates the profile line strategy successfully reduced high frequency variability, especially after 2021-08-21. Implementing the high-frequency averaging provides a better representation of the entire kiln circumference, which helps reduce the effects of rotational aliasing caused by non-uniform circumferential shell temperature variations (e.g., ring formation and partial refractory failure).

Rotational aliasing can lead to biased measurements which can result in the spurious high frequency shell temperature variations observed in Figure 4.20. This can in turn lead to false positives and high uncertainty of ring indications. In this

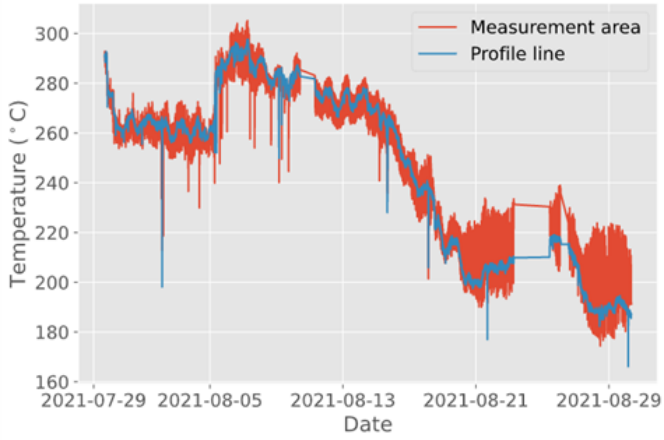


Figure 4.23: The profile line data with the highest correlation (blue) is plotted with the measurement area data (red) to demonstrate the improved noise reduction.

work a simple solution was proposed to improve the KST sampling procedure. This solution was implemented for one of the three thermal cameras and higher quality data was analyzed to confirm the rotational aliasing phenomena and validate the proposed solution.

Given the significant effort involved with manual validation it is recommended to improve the quality of available KST data in order to proceed with ring detection and diagnosis. Fortunately, the practical utility of the heatmap visualization method has yielded increased interest for kiln monitoring with thermal cameras. Moreover, the discovery of rotational aliasing provides industry guidelines such that future ring detection efforts can be undertaken with higher quality data. The final section of this case study addresses a separate data-driven kiln monitoring application, i.e., predicting residual calcium carbonate for kiln optimization.

4.4 Inferential Sensing for Residual Carbonate Prediction

Residual carbonate prediction refers to inferring the amount of calcium carbonate remaining in the calcium oxide nodules produced by the lime kiln. Predicting the

extent of endothermic calcination (described in reaction 4.1) based on the available operating variables enables kiln operators and/or model predictive control (MPC) algorithms to reduce specific energy and optimize kiln performance. Residual carbonate is one of the key target variables for controlling the quality of the lime product. This section provides background on RCC and previous residual carbonate literature before introducing the proposed methodology for hyper-inferential sensing. Experiments are conducted to develop an inferential sensor for residual carbonate prediction and demonstrate the utility of the experimental apparatus. Finally, this chapter concludes with a discussion of the results and recommendations for further development.

4.4.1 Residual carbonate control for kiln optimization

Traditional control strategies for rotary lime sludge kilns consist of independent PID control loops. The firing end temperature is controlled with the burner fuel flow subject to excess oxygen constraints. Excess oxygen and feed end temperature are controlled by manipulating the induced draft (ID) fan speed, subject to hood pressure constraints and ID fan limitations. Given the energy-intensity of lime kilns and their importance for white liquor production these units have been the subject of APC innovations. The inherent carbon-intensity of lime kilns is a further incentive that is becoming increasingly relevant.

The concept of using residual carbonate soft sensing to enhance lime kiln MPC has been around for at least two decades [23]. Lime kiln MPC has proven to be a successful approach for kiln control that can help minimize temperature fluctuations resulting in lower specific energy consumption, increased production, and higher quality lime product. Residual carbonate prediction is particularly important for reducing specific energy while maintaining sufficient product quality for downstream recausticizing processes.

Residual carbonate prediction

Residual calcium carbonate is a key product quality indicator and as such is an important target variable for kiln control. Product lime samples are collected by mill personnel once or twice per shift and subject to laboratory testing to determine the

residual carbonate content. The results from these lab tests are used by operators to provide key adjustments to kiln temperature setpoints. Developing an inferential sensor to predict residual carbonate can improve the resolution of these measurements to provide operators with better insights and/or to directly provide kiln temperature setpoints for closed-loop control. Shell temperature measurements from IR cameras have drawn interest to improve residual carbonate prediction.

The industrial collaborators involved in this research have successfully demonstrated improved residual carbonate prediction with thermal camera data [45]. Their work involved experiments with a first order state space model and three different sets of inputs, i.e., first principles energy balance models, traditional temperature measurements, and thermal IR camera measurements. A non-conventional regression metric referred to as 'fit quality' was used which is similar to the coefficient of determination (R^2), but it uses the ratio of the mean absolute error (MAE) to the mean absolute deviation (MAD), i.e.,

$$\text{Fit quality} = \left(1 - \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{|y_i - \bar{y}|}\right). \tag{4.8}$$

The method with the highest fit quality (26.4%) involved the thermal IR camera measurements, followed by the traditional temperature measurements (fit quality = 15.7%), and lastly the energy balance model (fit quality = 9.6%) [45].

Based on this work a closed-loop RCC strategy was implemented on an industrial lime kiln and shown to provide better product quality with reduced fuel consumption, leading to approximately $200,000 per year in fuel savings and emissions reductions of 3000 tonnes of CO_2 per year [44]. The contributions presented in this case study include building on previous literature to develop a robust residual carbonate prediction model while also introducing a hyperparameter optimization framework for automating soft sensor development. This framework is referred to as hyper-inferential sensing and it is discussed in what follows.

4.4.2 Hyper-inferential sensing

The steps involved in inferential sensor development are introduced in Chapter 2 and demonstrated in Chapter 3. Conducting a quantitative comparison of methods for inferential sensor development involves challenges such as data cleaning, feature engineering, and regression (or classification). Comparative experiments require consistent implementation with an emphasis on organization, documentation, and repeatability. Running experiments and tabulating results can become time-consuming and onerous for practitioners. Poorly constructed experimental frameworks lead to large amounts of technical debt which increases overhead, slows progress, increases human error, and produces less reliable results.

This work builds on experience from Chapter 3 to overcome these challenges by integrating the comparative experiments into a hyperparameter optimization framework. Hyperparameter optimization typically refers to optimizing the selection of user-specified model parameters. This work expands that scope to include broader design decisions involved with inferential sensor development. The result is a comprehensive experimental framework that applies an optimal sampling strategy to search a space of design decisions for feature engineering, regression method selection, and model hyperparameter selection. The proposed hyper-inferential sensing framework is demonstrated by providing a comprehensive comparison of methods for residual carbonate prediction.

The experiments conducted in this work are designed to build upon previous literature by providing a robust comparison of models with and without the thermal IR camera data. This experiment provides important insights to help industrial facilities understand the value of installing IR cameras for RCC. Note that the IR camera data studied in these experiments is not necessarily representative of KST measurement technologies generally. However, additional industrial case studies can be subject to this structured quantitative analysis to better understand the value of KST measurements for residual carbonate prediction.

The hyper-inferential sensing framework is extensible, allowing for convenient addition of new inferential sensing methods. It is designed to minimize the overhead associated with data preparation. Moreover, this framework is designed for a com-

mon industrial application, i.e., soft sensing of key operating parameters that are measured at an irregular frequency with analyzers or laboratory tests. By following simple input formatting instructions it can be applied to develop soft sensors for other key laboratory measurements in various industrial processes. Key elements of the hyper-inferential sensing framework are introduced before presenting the experimental results.

Problem formulation and data structuring

Previous literature on residual carbonate prediction involved experiments with a first order state-space model. This formulates the prediction problem as a time-series forecasting method, which is common for applications related to dynamic control. However, the hyper-inferential sensing presented in this work formulates the prediction problem as a traditional supervised learning method. This reformulation is motivated by the nature of the data for this type of application.

The target variable is a laboratory measurement with an irregular frequency that is measured relatively sparsely. The features are generally PVs that are measured with regular, high frequency sampling intervals. Therefore, implementing predictions with a first order state-space model requires resampling such that the target and the features have the same sampling frequency. This resampling can significantly distort the training data and the resulting predictions. Advanced imputation methods can provide better results than naive imputation. Continuous time series methods such as neural ordinary differential equations are a promising research direction for this application. However, in this work a straightforward approach is used to leverage the time-series nature of the data while maintaining the practical benefits of the traditional supervised learning paradigm.

Data structuring involves identifying samples when the target variable changes, capturing windows of PVs, flattening the PVs into lagged feature columns, and adding features for lagged target values and the periods between subsequent samples. Invalid data are simply replaced with NAN values which are imputed with a zero order hold interpolation. Therefore, the resulting data may contain invalid samples (e.g., sensor faults) represented as flattened features with zero variabil-

ity. This simple data cleaning strategy does not leverage domain knowledge, but it does enable the hyper-inferential sensing framework to scale to many features and transfer to other applications. To address the presence of potential invalid samples an outlier removal method is used.

Outlier removal

The fidelity of industrial process data should not be taken for granted. Manual investigation of data with applied domain knowledge is the safest approach to identify invalid samples. However, this method is onerous and does not scale efficiently. The purpose of the proposed experimental framework is to provide an efficient and reliable tool for automating inferential sensor design decisions with quantitative analysis of industrial process data. This can supplement the performance improvements that can be obtained through more rigorous data preparation.

The primary role of outlier removal in this work is to compensate for the minimalist approach to data preparation. Minimal data preparation enables the hyper-inferential sensing framework to be more generally applicable to many soft sensor applications. However, upset process conditions, shutdowns, sensor faults, and programming errors can corrupt industrial process data and produce invalid samples. Therefore, a small number of outlier samples are removed to improve the quality of the data for modeling.

Uniform manifold approximation and projection (UMAP) is used for outlier removal in this work. Theoretically, UMAP is a contemporary approach for dimensionality reduction that is based in Riemannian geometry and algebraic topology, details of which are beyond the scope of this work [84]. Practically, UMAP is a dimensionality reduction technique that excels at preserving both the local and global structure of complex datasets. The process analytics community has studied UMAP for visualization, dimensionality reduction, and outlier removal [56].

The use of UMAP in this work is primarily for clustering to remove samples that are suspected to be invalid. Clustering is applied by specifying the number of local neighbors and the size (or number of components) of the lower dimensional encoding. A threshold is also specified that controls the amount of outlier samples that

are removed. A conservative approach is taken to remove only 6% of the training data. The UMAP clustering for outlier removal of the training data is demonstrated in Figure 4.24. In addition to removing outliers, UMAP is valued for visualiz-

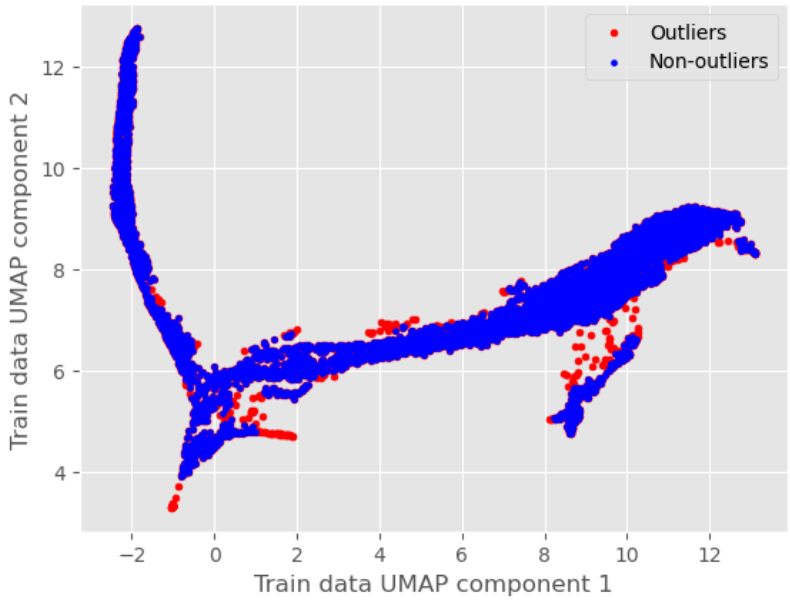


Figure 4.24: Clustering with UMAP to visualize the samples of high dimensional industrial process data that are used to train the inferential sensor models (blue). Outlier samples (red) are discarded for modeling but they can be investigated to yield operating insights.

ing clusters of process data and investigating the clusters while applying domain knowledge. Outlier removal is a source for further improvement of the experimental setup. Additional methods should be studied and the selection and specification of these methods should be included within the scope of the experimental trials.

Experimental trials

The experimental trials begin after outlier removal and they involve sequentially learning better inferential sensor designs with respect to the choices for feature learning, model selection, and model hyperparameter selection. Figure 4.25 illustrates the experimental setup for the hyper-inferential sensing strategy. The feed-

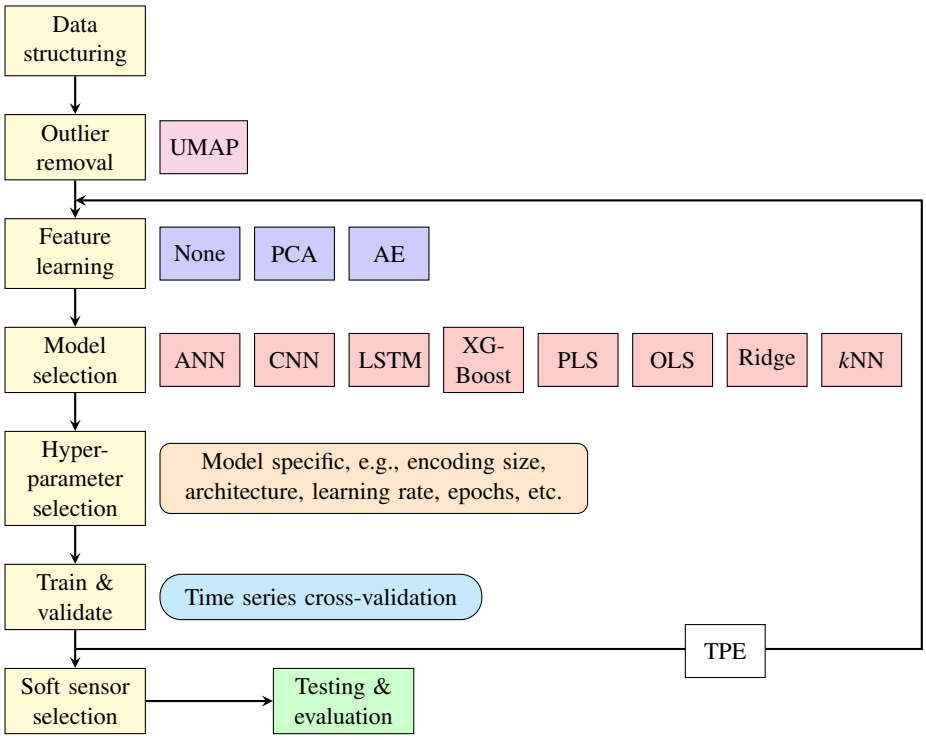


Figure 4.25: Experimental setup for hyper-inferential sensing. A variety of traditional and contemporary methods for process analytics and ML are selected to provide a highly structured framework for efficient and reliable quantitative comparisons.

back of the validation score to the input of the feature learning stage represents the scope of the experimental trials. In this work the validation score is the average mean squared error from time series cross-validation experiments with the current trial model. The trial model with the best cross-validation performance is then selected and evaluated with previously unseen test data.

Bayesian optimization with a TPE is used to guide the experimental trials by selecting and preparing the candidate soft sensor [6]. The TPE algorithm learns from past experimental configurations to select actions that maximize the expected improvement of the current trial model. The available choices for feature learning and model selection are presented in Figure 4.25. Implementation and hyperparameter

specification of these methods is described in what follows.

Features, models, and hyperparameters

The majority of methods for feature learning and regression in the proposed experimental framework have already been introduced. The space of available methods in Figure 4.25 provides a variety of traditional methods (e.g., PCA and PLS) and advanced deep learning architectures (e.g., AE, ANN, CNN, and LSTM). Studying both proven and innovative modeling methods with distinct structures provides both flexibility and insight into beneficial methods for specific applications. This can accelerate soft sensor development by helping researchers identify promising avenues for further investigation and fine-tuning. A brief discussion of the new methods, namely ordinary least squares (OLS) linear regression, ridge regression, *k*NN, and extreme gradient boosting (XGBoost) is provided here along with details on model implementation and hyperparameter selection.

Many industrial soft sensors use OLS linear regression models for convenience and interpretability. Coefficients are fit to linear features such that the sum of the squared errors between the training targets and predictions is minimized. The ridge regression model extends OLS to include ℓ_2-norm regularization of the coefficients. The OLS model does not have hyperparameters, but ridge regression has a weight for the ℓ_2 regularization that can vary between 0.0001 and 1 in this study.

The *k*NN regressor is a non-parametric method that generates predictions by finding the k most similar training samples and averaging their targets. Similarity in this work is measured by the Euclidean distance between the high-dimensional feature space of samples. Alternatively, XGBoost is an ensemble ML algorithm that uses a concept known as boosting which involves using the ensemble of models sequentially to predict the residuals of prior models. The outputs of the models are combined to generate predictions. Gradient boosting uses gradient descent to optimize the loss function while training the ensemble [97].

The space of available hyperparameters depends on the selected model. Deep learning models share common hyperparameters such as number of epochs, learning rate, batch size, number of hidden layers, and hidden layer size, among oth-

ers. Generally, the deep learning soft sensor configurations have a significantly larger space of potential hyperparameters and require a significantly larger amount of computation effort per experimental trial. The specific hyperparameter search space for each model is described in Appendix A, with traditional ML hyperparameters listed in Table A.6 and deep learning hyperparameters listed in Table A.7.

4.4.3 Experimental results

Over sixty different kiln operating variables are used to demonstrate the hyper-inferential sensing framework and develop a residual carbonate soft sensor. The introduction of each of these variables is beyond the scope of this work but a brief description is provided in Table A.5.

The hyper-inferential sensing framework is used to develop two residual carbonate soft sensors, i.e., one with IR camera data, and one without IR camera data. Each of the studies involves 1000 experimental trials, where each trial represents the configuration, training, and validation of the proposed inferential sensor. The trial results from both studies are saved and combined to analyze the validation results of various experimental setups. Across all 2000 trials, the average validation scores of the feature learning methods are presented in Figure 4.26 and the average validation scores of the regression methods are presented in Figure 4.27. The pie charts in Figure 4.26 and Figure 4.27 represent the proportion of the 2000 trials that are dedicated to the associated method. To reduce the influence of outliers, the validation errors are clipped at a mean squared error (MSE) of 100. Recall that the MSE of each trial is the average MSE of three time-series based validation experiments.

On average, the auto-encoder feature learning method provides the lowest validation error. The *k*NN, LSTM, and XGBoost regression methods provide the lowest average validation error. However, Figure 4.26 and Figure 4.27 do not tell the whole story. Even with clipping, the average validation score across all 2000 trials is dominated by the unusually bad trials, that are not particularly relevant. Instead, the best performing methods are of interest. The XGBoost regression method with the raw features is the best performing residual carbonate inferential

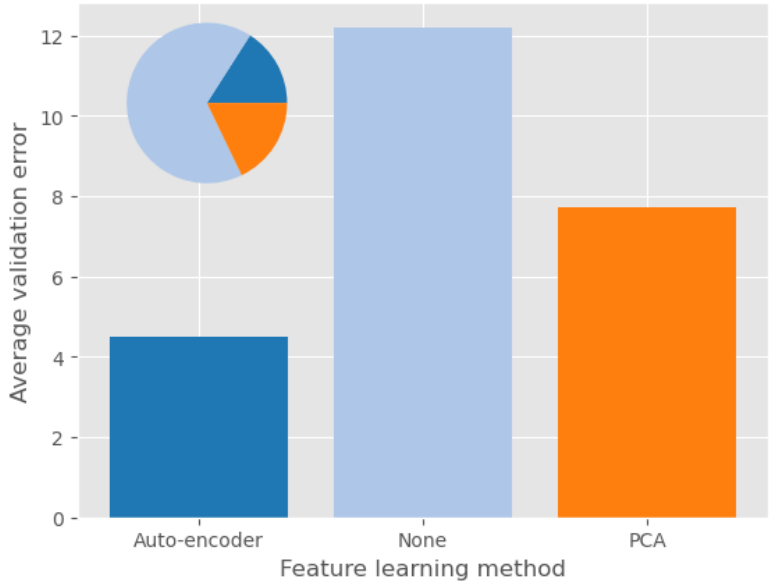


Figure 4.26: Average validation error of the feature learning methods across 2000 experimental trials. The proportion of trials dedicated to each method is depicted in the pie chart

sensing method for both studies (i.e., with and without IR camera data). Slightly different hyperparameter specifications are learned for each study.

The best experimental trial from each study is used to configure candidate residual carbonate soft sensors with and without IR camera data. These soft sensors are re-trained on the entire training dataset and used to predict previously unseen testing data. Given that model robustness is valued in this investigation, the soft sensors are tested on roughly 3.5 months of operation without any feedback, updating, or intervention. The test-set predictions of the soft sensors are compared with the true laboratory measurements in Figure 4.28. The evaluation shown in Figure 4.28 demonstrates the enhanced ability of the soft sensor with IR camera data to capture the variability in the measured residual carbonate content.

As Table 4.2 demonstrates, the IR camera data significantly improves the performance metrics of the residual carbonate predictions. The coefficient of determination (R^2), fit quality, and RMSE are all improved by the use of IR camera data.

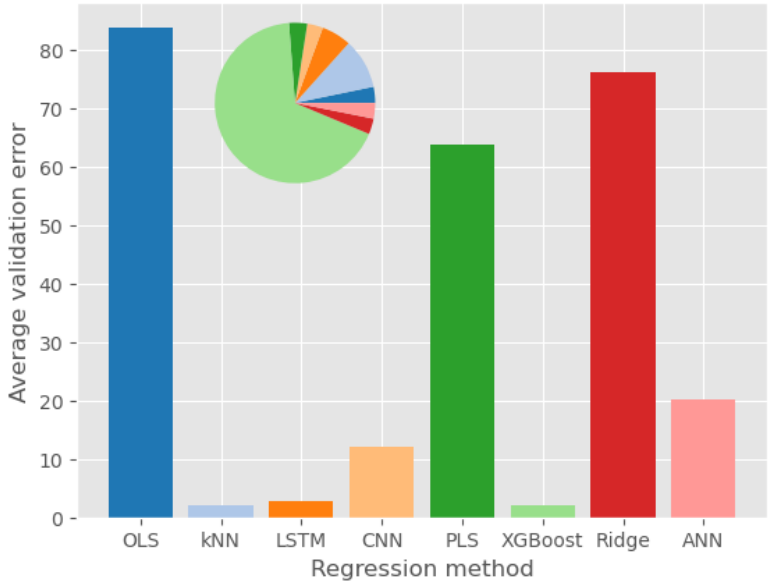


Figure 4.27: Average validation error of the regression methods across 2000 experimental trials. The proportion of trials for each method is displayed in the pie chart

Table 4.2: Summary of results from testing residual carbonate predictions developed with the proposed hyper-inferential sensing framework.

Inferential sensor	R^2	Fit quality	RMSE
With IR camera data	0.383	0.267	1.36
Without IR camera data	0.225	0.174	1.52

This result is consistent with the principles of lime production and the results of previous literature.

4.4.4 Discussion and recommendations

This study enhances the evidence that KST data enhances residual carbonate prediction with robust experimentation and an alternative problem formulation that involves minimal imputation. There are various reasons that explain the mediocre quality of the predictions. The target variable has significant uncertainty due to

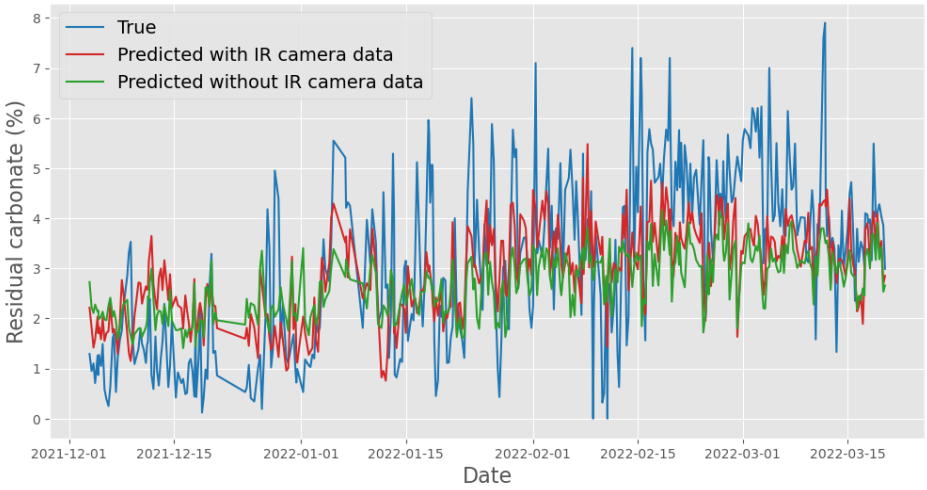


Figure 4.28: Testing the candidate residual carbonate soft sensors developed with IR camera data (red) and without IR camera data (green).

the nature of the laboratory measurement. The quality of the IR data has already been demonstrated to have shortcomings. These issues may not have been sufficiently addressed by outlier removal, but the IR camera data still provides significant value. Finally, residual carbonate prediction is an inherently challenging problem with both fast and slow dynamics. This is evidenced by the eerily similar fit quality of previous literature (26.4%) to the value obtained here (26.7%) despite the fundamental differences in problem formulation.

The high proportion of trials spent on the XGBoost model is driven by their low average validation error. The TPE algorithm searches promising regions based on past results. These regions were often XGBoost models with the raw features. The results presented here are obtained by simply evaluating the best experimental setup on roughly 3.5 months of new operating data. Additional refinements such as re-training and/or adding a model bias can lead to further improvements.

The hyper-inferential sensing framework is already comprehensive, but it can be improved in many ways. Additional methods for outlier removal should be studied and more rigorous manual investigation should be applied to compare classes. The breadth of the experimental trials should be increased to include selection and

specification of the outlier removal method. Increasing the exploration and revising the available search space is recommended to provide more insightful experiments. More extensive testing should be conducted to provide insights into model generalizability. The hyper-inferential sensing framework should be applied to different regression problems in different industrial facilities, and the regression method should be extended to address classification and time-series forecasting problems.

Chapter 5

Pitfalls and Guidelines for Industrial Process Analytics

In Chapter 2 a high-level procedure for learning from industrial process data is presented along with challenges that are commonly encountered by researchers and practitioners developing data-driven APC solutions. This chapter reflects on experience obtained through literature, case studies, and industrial practice to provide guidelines that help practitioners overcome these challenges. Some of the material from this section has been published as follows:

- S. C. Lim, S. Elnawawi, L. D. Rippon, D. L. O'Connor, and R. B. Gopaluni. Data quality over quantity: Pitfalls and guidelines for process analytics. *IFAC World Congress 2023*, pages 1–8, 2023.

5.1 Identifying and Framing a Data-driven Opportunity

Discovering novel applications that can realistically achieve significant operating impact represents a significant contribution to advancing the efficiency and sustainability of resource-intensive manufacturing. With limited resources, prioritizing which opportunities to pursue is important. Ultimately, the primary challenges associated with identifying and framing a data-driven opportunity involve ensuring

sufficient impact and confirming feasibility of the proposed solution.

Impact and benefit estimation

Estimating benefits, such as reduced fuel consumption, can provide important motivation and justify expenditure of resources on innovative solutions. Benefits can be quantified in terms of safety incidents, environmental impact, and/or economic objectives (e.g., quality and productivity). Site personnel can provide guidance to identify data-driven opportunities with significant impact. Neglecting the importance of justifying the research motivation is a common pitfall that can lead to loss of focus, lack of cooperation, and loss of funding.

Obtaining accurate quantitative estimates is challenging prior to data acquisition. Simple estimates based on key process data can provide meaningful insights. It is important to consult the facility personnel as they may have domain knowledge to guide and simplify benefit estimation. In the absence of historical process data, potential benefits can be roughly quantified with minimal *a priori* process knowledge (e.g., the scale and efficiency of the unit operation, and the frequency of the fault).

Large-scale industrial processes with recurring faults that disrupt key energy intensive unit operations are often implicitly understood to justify research efforts. Especially for smaller projects that do not require justification of significant expenditure. In some cases, the impact is known to be significant enough for industrial collaborators to actively seek innovative solutions (e.g., the arc loss fault). In other cases, literature and SMEs can be consulted because the fault is already notorious (e.g., ring formation in lime kilns).

Alignment on objectives and resource requirements

Stakeholders should align on well-defined research objectives to avoid competing interests that distract from the core objectives. These objectives should be realistically achievable given the available resources and stakeholder participation. Vetting the quality and reliability of the available resources is critical for ensuring project feasibility. Potential resources include instrumentation, literature, documentation (e.g., P&IDs, control narratives, etc.), remote access, and the availabil-

ity of site experts. Determining the measurement quality and failure modes of key instrumentation can reveal insights related to project feasibility.

Availability and integrity of a source of ground truth

Given raw industrial data, it is important to consider whether there is a practical means of generating reliable labeled outputs from this data. Can the data be labeled automatically, e.g., with an algorithmic definition of the fault? If not, can the data be labeled manually or is the raw data too ambiguous? Are stakeholders willing to undertake the potentially cumbersome task of manual data labeling? Solutions that are based on supervised ML methods need to carefully consider the availability and integrity of a source of ground truth.

Some applications (e.g., soft sensors) may have ground truth data available in sparse, irregular frequency lab samples. If a source of ground truth is not available then a reliable manner for identifying the ground truth labels should be defined. If labels cannot be identified in a reliable manner then stakeholders need to realign on the objectives, proposed solutions, and resource requirements, because accurate labels are a pre-requisite for supervised learning.

In the pyrometallurgy case study, labeled data is not provided. Therefore one of the initial tasks is to generate accurate labels. This task is enabled by consulting industrial collaborators and developing an algorithmic definition of arc loss that can be reliably applied to the historical data. The label-friendly nature of the pyrometallurgy case study makes it an ideal case study to use for an FDD benchmark challenge.

During the ring formation case study, a succinct algorithmic definition is not available. Given the limitations on the available data, efforts towards validating ring detection suffer from inherent ambiguity and subjectivity. Therefore, the objective is realigned to investigate and address the unexplained source of the validation uncertainty (i.e., rotational aliasing). For residual carbonate prediction, lab samples are used as a source of ground truth, which is common for inferential sensors.

5.2 Data Acquisition and Contextualization

Acquisition and contextualization of industrial operating data is best performed with an emphasis on process knowledge and in collaboration with process experts. The following guidelines are provided to help practitioners avoid common pitfalls specific to applications with industrial process data.

Establishing connections with data sources

A key challenge that limits the progress of collaborations between academia and industry is obtaining sufficient access and connectivity to explore the available data sources, assess their integrity, determine the data requirements, and review important contextual information. Establishing a secure and reliable connection to the required data sources is essential for productive collaborations and efficient solution development. Developing a novel data-driven solution often requires iterative scope refinement and experimenting with alternative data sources.

Relying on email interactions with site personnel to transfer data is a fragile and inefficient strategy that can create significant delays and inhibit research progress. Site personnel are often preoccupied with higher priorities such as ensuring safe and reliable production. Data requests can take weeks to fulfill, and in some cases may be forgotten altogether. Employee turnover and insufficient oversight can result in periods without access to critical data sources.

Acquiring useful industrial data often requires an iterative exercise with collaborative efforts between researchers and plant personnel. Therefore, it is much more productive to provide direct historian access (e.g., remote access) instead of manually sharing data. However, convenient access to data sources is not always practical due to company policies and/or insufficient digital plumbing. Many processes are just beginning their digitalization journey which has been a large impediment to learning from industrial data. As modern OT is adopted, data connectivity will significantly improve, enabling much more convenient and immersive research into advanced manufacturing facilities.

Selecting the industrial data

Given the historical difficulty of acquiring industrial data, there is an understandable tendency to initially maximize the volume and variety of data acquired. However, if a reliable connection is readily available then it is often better to start with a smaller (more manageable) number of PVs that are core to the investigation. Large, complicated datasets can slow progress by creating a high barrier of domain knowledge and by incurring technical debt for data loading, processing, and exploration. Studying core PVs that are highly relevant to the scope and then adding complexity in small, coherent batches is an effective strategy for exploring data while developing and applying process knowledge.

Domain expertise is critical for understanding the available data and selecting relevant variables from suitable periods of operation. For example, control modes may be label encoded which can be corrupted by time averaging. The research objectives should guide the resolution and other extraction settings (e.g., interpolation) of the acquired data. Consider the effect of time averaging on upsets and high frequency process dynamics then reconcile these effects with the desired research objectives. Linear interpolation may not be appropriate for sparsely measured variables such as lab measurements. First principles process knowledge can help identify key parameters and eliminate irrelevant variables. Ultimately, data exploration and additional contextualization is necessary to select relevant variables and sample suitable periods of operation.

Exploratory analysis and data visualization

In process analytics, exploratory data analysis (EDA) is essential. Domain knowledge is essential for inspecting and understanding the data during EDA. Many practitioners under-appreciate the role of domain-specific data visualization for effective collaborations. Developing novel visualization methods empowers operators and SMEs to conveniently apply their domain expertise to the data. This is a proven approach to improving operating outcomes. The visualization strategy presented in Chapter 4 has already been applied to multiple mills to troubleshoot ring formation and improve operation [30, 103].

Simple charts such as line plots, histograms, and box-plots can be used to manually investigate PVs. Clustering algorithms (e.g., UMAP) can help guide data exploration and yield novel insights. Control narratives and additional documentation can be used to help interpret the data and provide further direction for exploration.

Data contextualization

As Figure 5.1 demonstrates, process data is just one of many different sources of valuable information in an industrial process. Plant topology data from process & instrumentation diagrams (P&IDs) provides structural context, informing the practitioner on causality information. Maintenance records can provide important context such as fouling, equipment degradation, and instrumentation reliability to determine the integrity of the measurements. Control narratives are critical for interpreting and evaluating the behavior of complex control systems. Lab quality data provides key operating parameters that are often useful for soft sensor development. Insights into abnormal events can be obtained from alarm data such as bad actors, alarm floods, and various metrics from the International Society of Automation (ISA) 18.2 standard [49]. Each of these different types of data can provide valuable context for learning from historical process data.

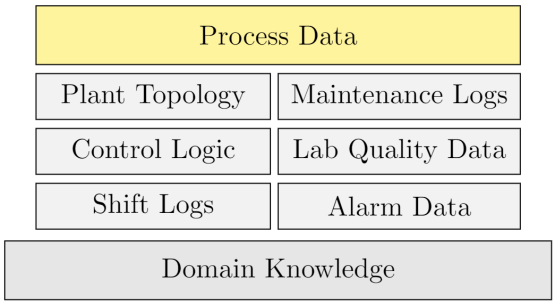


Figure 5.1: Supplementary sources of information are necessary to contextualize process data and obtain meaningful insights [76].

Convenient read-only remote access to the process historian, DCS graphics, and the other data sources described in Figure 5.1 can help reduce the burden on mill resources. However, there is no substitute for the years of industrial experience possessed by mill personnel. Therefore, access to experienced site contacts is an-

other requirement for properly contextualizing industrial process data. During the ring formation case study mill personnel are consulted to provide valuable context to periods of suspected fouling with the shell temperature heatmap and relevant process data.

Shortcomings of industrial data

It is important for data practitioners to question the integrity of not only the measured values, but also the metadata, e.g., units of measurement, variable descriptions, and tag names. How do practitioners ensure that the measurements accurately represent reality or that the tag descriptions and/or units of measurement are correct? These are practical challenges that can have a significant influence on research outcomes if they are not considered during data acquisition. Awareness of the shortcomings of both measured and calculated variables is critical for developing robust solutions.

The IR camera obstruction in Figure 4.17 and the discovery of rotational aliasing demonstrate the importance of interrogating the quality of the industrial process data. Maintenance of instrumentation to produce high quality data in large-scale industrial processes is a significant challenge. Practical constraints and human error can result in sub-optimal installation, configuration, and maintenance. One action that can be taken is for researchers and/or industrial collaborators to carefully review the data that is used for process analytics. This can involve researching best measurement practices, consulting with instrumentation specialists, and conducting field inspections. Early flagging of data shortcomings can prevent costly downstream troubleshooting efforts.

5.3 Data Preparation

Although data pre-processing is often unfairly maligned as trivial and technically uninteresting, in practice it has an out-sized influence on the success of real-world AI applications. Industrial data must be prepared in a manner appropriate for both the process opportunity and the proposed solution. Despite the importance of data preparation, there is relatively sparse guidance in the literature on how to handle

process data for analytics and ML.

Data validation and reconciliation

Problematic data is not always avoided during data acquisition. Therefore, additional efforts must be taken to validate and reconcile the integrity of the raw data. Invalid process data can occur from poor instrument installation, lack of sensor calibration, fouling, and many other reasons. Data reconciliation can be performed by using redundant measurements, mass and energy balances, data visualization, domain knowledge, and operator insights. Consulting with SMEs and site personnel to obtain historical accounts of the measurement quality can provide valuable insights. Understanding the source of the measurement (i.e., instrument type, laboratory procedure, or calculation) allows practitioners to understand the inherent limitations of the data.

For key variables, simple linear regression models can be designed with process knowledge, trained on suitable data, and applied to produce predictions. Prediction residuals can be analyzed to identify periods with suspicious measurements. A network of such models can be applied to monitor key measurements across the plant (e.g., major chemical, energy, and material inputs and products). The residuals can be monitored to systematically identify process upsets and potentially invalid data.

Data structuring

Ultimately, the desired problem formulation dictates the data structuring procedure. If the data is sampled at regular intervals with a common date-time index then the problem can be conveniently formatted as a multivariable time series forecasting problem. For densely sampled variables, small amounts of missing data can be imputed with simple interpolation methods. Sparsely sampled laboratory data that is obtained at irregular intervals cannot be imputed as easily. Although these details are essential for experimental reproducibility, they are often not presented in a thorough and comprehensive manner in literature.

Data structuring is demonstrated during the development of the arc loss and residual carbonate inferential sensors. The residual carbonate target samples are identi-

fied by selecting value changes because the data is acquired at five minute intervals as zero-order hold interpolations from the lab entry. Features are constructed as lagged values of the available operating variables from windows that are specified to capture the relevant process dynamics. A similar sampling procedure is conducted for the pyrometallurgy case study to construct a balanced dataset of arc loss faults. Lagged target variables can be included as an additional feature to emulate auto-regression. The time difference between subsequent samples can be encoded as a Unix epoch to provide an indication of duration.

Data cleaning

Data cleaning involves analyzing the structured dataset and applying changes that minimally distort the data while making it amenable for modeling. This includes addressing artefacts and inconsistencies from the DCS. Different PVs in the same dataset can be configured differently with respect to the values produced during shutdowns. For example, NAN values might occur for one variable, whereas another PV might have all zeros during a shutdown. Some variables might have strings encoded during a shutdown such as 'Bad Input' or 'Bad Status', while others may have a linear (or zero-order hold) interpolation.

Data cleaning at scale, with many samples and a large set of variables can be a significant challenge. Studying the modes of operation can provide valuable insights. Figure 5.2 demonstrates an archetypal example of a PV partitioned into four common operating modes. Figure 5.2 shows a histogram of combustion air flow rates and Figure 5.3 shows sampled periods of time series data corresponding to each mode of operation. Many PVs share a highly similar distribution which can help simplify data cleaning. However, this insight cannot be applied blindly as some PVs can have significantly different distributions (e.g., excess oxygen). Normal operating ranges for key operating variables can be obtained by consulting plant personnel.

The mode of key control loops should be considered when preparing process data. Check that the regulatory control system is functioning well and the control loops are well-tuned. If the control loops are oscillating, the data can be detrimental

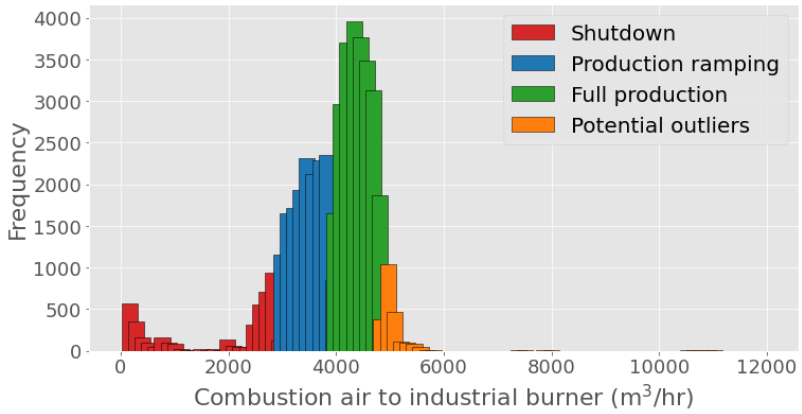


Figure 5.2: Histogram of process data with different modes of operation.

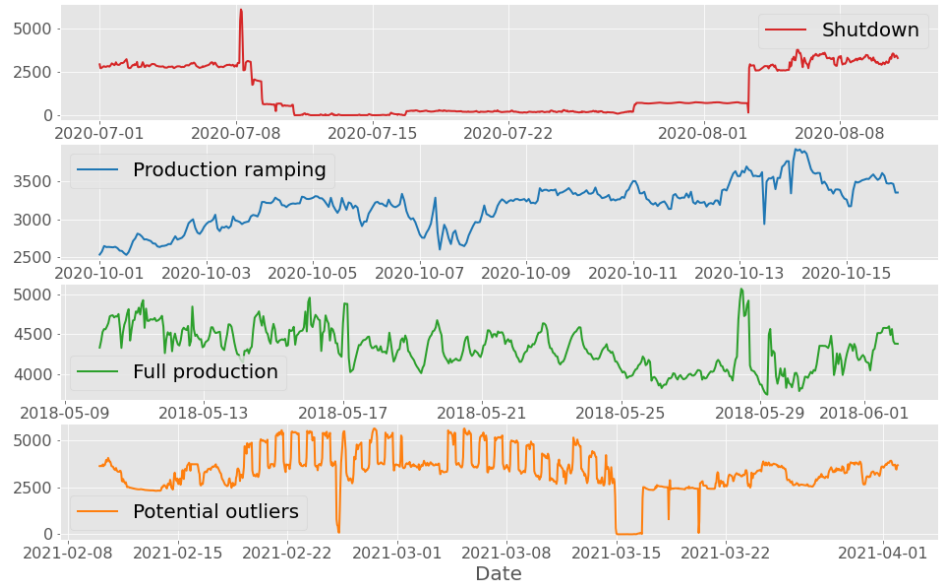


Figure 5.3: Visualizing data from different modes of operation.

for model performance. Tags that are indicative of different operating conditions should be identified. For instance, in pulp mills, valve positions can indicate the use of different process fluids (e.g., filtrate, mill water, condensate, or white water). Samples collected when the process is not at steady-state may reflect transient process conditions that may not be suitable for the desired modeling objectives.

Feature engineering and representation learning

Feature engineering and representation learning are often considered optional aspects of ML workflows. Some models perform representation learning implicitly (e.g., CNNs). Feature engineering involves manually transforming the feature space by applying domain knowledge. For example, calculating an important operating parameter (e.g., specific energy) and adding it to the set of features can potentially improve model performance. Representation learning (or feature learning) is distinguished for automatically generating new features with dimensionality reduction methods.

In this work the feature learning methods that are studied do not provide significant performance benefits for predicting either arc loss or residual calcium carbonate. For residual carbonate prediction the AE method is shown to provide the lowest average validation error, but ultimately the best model performance was with the raw features. Nevertheless, refining the feature space of the input data can help improve model robustness and reduce the computational burden, which can in turn enable a more rigorous search of model hyperparameters.

5.4 Method Selection, Development, and Evaluation

Development and evaluation of modeling methods receives significant attention in literature, but practical considerations and best practices for working with industrial process data are rarely addressed.

Storing and loading industrial process data

Industrial operating data is often confidential which can preclude the use of convenient third-party cloud-hosted environments (e.g. Google Colab) to store and load large quantities of process data for subsequent analysis. without these services, analysis of large amounts of data can strain the available computational resources and slow progress. Storing large quantities of data in Excel files is a common source of inefficiency as these Excel files can take orders of magnitude longer to load into the programming environment than alternative methods. To save time and overcome memory limitations, practitioners can store their data as `.parquet`

files instead of spreadsheets. Another approach to saving memory is to set the data types of each variable (e.g., float-32 instead of float-64).

Selecting suitable methods

Method selection can be informed in many ways including with quantitative experiments and by investigating the theoretical properties of the available algorithms [123]. A 2010 survey revealed that over 90% of industrial soft sensors in Japan use linear modeling techniques like OLS and PLS [58]. Fundamental problems regarding the misalignment between methods proposed in literature and those used in practice need to be addressed.

Despite the increasing sophistication of ML tools in recent years, the implementation of such algorithms is limited by the hardware capabilities in the plant. Depending on the digital sophistication of the facility, complex models like DNNs may not be the most appropriate or even feasible solution. Feasible models should be considered first if they help drive business value.

Selecting suitable methods needs to be reconciled with the objectives of the particular application. In some cases (e.g., residual carbonate prediction), the quality of the predictions is paramount, and black-box models that drive residuals to zero can be appropriate. In other cases (e.g., ring indication), the residuals are used as indicators so it is more appropriate to select models with results that are easier to interpret. The selection of suitable methods can also be determined by the problem formulation, which can be dictated by the state of the available data.

Metrics that align with outcomes

For industrial applications, striving endlessly towards better performance metrics such as MAE, RMSE, or coefficient of determination (R^2), can be time-consuming and counter-productive. It is important to consider the business impact and intended usage of the models. For example, exogenous regressors can be variables like the flow of fuel to the burner or the flow of lime mud to the feed-end of the kiln. A common pitfall in process monitoring applications of ML is that domain knowledge is disregarded and selection of exogenous regressors is reduced to a

data-driven endeavour to optimize a performance metric. In practice, it is often beneficial to select exogenous regressors with careful regard for the operating behaviour they help represent.

The presence of sensor noise, measurement errors, uncertainties in lab results, and other practical considerations may significantly outweigh minor performance improvements in inferential model metrics. The production losses incurred by an inferential model being unreliable or unavailable may be several orders of magnitude greater than the cost of a model with a slightly lower metric. Model metrics like the prediction error are only a small part of delivering value in practice.

Conducting comprehensive experiments

Although model metrics are not sufficient for developing industrial solutions, they are nevertheless important for evaluating potential ML models. Conducting comprehensive comparative experiments with different modeling configurations is a challenging endeavour that consumes large amounts of research capacity. As the hyper-inferential sensing study demonstrates, it is critical to develop structured experimental frameworks to conduct productive experimental comparisons.

Eliminating technical debt is a key aspect of successfully conducting comparative experiments. Creating modular, extensible model development frameworks such as the one demonstrated in Figure 4.25 are necessary to efficiently conduct the complex quantitative experiments. Relying on manual intervention to configure experiments and tabulate intermediate results can result in human-error and poor experimental reproducibility.

5.5 Deploying Sustainable Solutions

Successfully deploying solutions and obtaining sustained utilization is critical for achieving long-term operating benefits. Key considerations are presented here to help process analytics researchers deploy sustainable solutions.

Understanding trends in utilization

When APC solutions are not trusted by operators, they are often disabled or ignored. Utilization refers to the percentage of relevant operating time that the data-driven solution is active. Once a solution is deployed it is important to monitor utilization and investigate factors that lead to reduced utilization. Enhancing the proposed solution to address these factors can be critical for the long term success of the proposed solution.

Many different factors can contribute to poor utilization. The process can drift over time due to changes such as equipment conditions and feed quality. The proposed solution may not be designed to sufficiently handle inevitable upset conditions. Instrumentation issues such as probe fouling and plugged pressure taps can result in invalid measurements. Laboratory measurement quality can vary depending on the experience of the technician and the reliability of the apparatus.

If possible, initially deploying the solution as a monitoring method can help gather data on potential shortcomings, guide revisions, and gain trust before a closed-loop solution is implemented. Awareness of the various process conditions that can lead to performance degradation of the proposed solution is key for developing sustainable solutions.

Robustness, durability, and model maintenance

How do we design durable soft sensors that maintain high accuracy for long periods of time given the dynamic nature of the underlying process? Model durability is the length of time a model is in production before performance degradation necessitates tuning or retraining. A high-performing model may not be a durable model. A durable model, even with a slightly lower performance relative to a fragile model, requires less frequent maintenance and would typically be preferred by plant personnel.

With respect to model development, experiments should be designed with consideration for the constraints of the industrial environment. The most pressing problem with industrial inferential sensors is not model accuracy, but rather, model maintenance. During experiments, it is easy enough to re-train a ML model on new data.

However, frequent model updates are not necessarily straightforward in industrial environments.

While developing the residual carbonate inferential sensors roughly 3.5 months of previously unseen test data is used for evaluation. Instead of updating the candidate model as new samples are processed, the original model is applied to all of the test data. This may seem unnecessarily difficult, but it is done intentionally as the long-term robustness of the predictions is of more interest than minimizing the prediction error.

Appreciating end-user experience

Neglecting the experience of end-users is a common pitfall that limits the impact of many proposed solutions in process analytics. From an operational perspective, the practitioner should understand how the solutions they are developing will be used to drive plant improvements. Unlike other domains, many of the models deployed in safety-critical industrial processes must be interpretable for plant personnel to understand, trust and maintain them. Utilization is required for impact and operator buy-in is key for utilization. Introducing robust solutions that are transparent and easy to interpret is essential for obtaining operator buy-in to drive sustainable utilization.

Failure to consider the end-user experience leads to solutions that are not adopted in practice. Minimizing unnecessary technical complexity in the presented solution can improve the end-user experience. Training operators to use and understand the solution can provide valuable feedback that can then be integrated to improve the operator experience.

Given the diverse nature of potential industrial applications and the variety of industrial data types it is impossible to prescribe an optimal data workflow for leveraging industrial process data. Instead, this chapter presents practical guidelines and common pitfalls encountered by researchers and practitioners while learning from industrial process data.

Chapter 6

Conclusion

This dissertation addresses an increasingly important societal problem, i.e., learning from large volumes of historical operating data to enhance the safety, reliability, and performance of large-scale industrial processes. Proprietary operating data can be shared and AI can be applied to learn best practices and maximize the sustainability of natural resource development in Canada and beyond.

In addition to providing practical guidelines for applying process analytics and ML to industrial operating data, this dissertation presents two distinct case studies. Specific contributions are presented for each industrial case study. The unifying theme of process analytics application with industrial process data provides insights into key challenges and practical guidelines for researchers and practitioners. A brief summary of the various contributions discussed in this dissertation is provided as follows:

- **Contributions from the pyrometallurgy case study:**
  - Introducing a novel FDD problem and preparing industrial data to formulate arc loss as a supervised ML classification problem.
  - Developing a novel inferential sensor model to predict arc loss.
  - Comparing traditional and contemporary process analytics methods for

representation learning in the context of arc loss prediction.

  - Presenting *The Arc Loss Challenge*, a novel FDD benchmark for ML methods with open-source industrial operating data.

- **Contributions from the kraft pulping case study:**

  - Introducing a novel kiln shell temperature visualization strategy that has demonstrated commercial success.

  - Discovering the rotational aliasing phenomenon.

  - Improving the KST measurement strategy to address rotational aliasing.

  - Introducing a novel hyper-inferential sensing framework.

  - Developing a residual carbonate soft sensor.

  - Confirming the value of IR cameras for residual calcium carbonate prediction.

- **Contributions for learning from industrial process data:**

  - Introducing use-cases for industrial process data and proposing a data-driven APC solution development strategy.

  - Presenting a comprehensive set of practical challenges encountered in process analytics applications.

  - Leveraging experience from case studies to provide guidelines and best practices for learning from industrial process data.

Given the breadth of this research, there are many areas for improvement and future work. The arc loss challenge has been developed and recently published. Additional efforts are required to investigate ring formation with high quality data that can resolve the ambiguity of validating ring indications. The hyper-inferential sensing framework can be improved in many ways and applied to different applications to compare soft sensor designs and provide insights into the value provided by supplementary instrumentation (e.g., IR cameras).

Bibliography

[1] Dall-e 2 by openai. https://openai.com/dall-e-2/. April 6, 2022 — Research, Multimodal. → pages 10, 31, 62

[2] Michael I Jordan: Machine learning, recommender systems, and future of AI — Lex Fridman podcast #74. https://youtu.be/EYIKy_FM9x0. Feb. 24, 2020. → page xxiv

[3] E. Acuña and C. Rodriguez. On detection of outliers and their effect in supervised classification. *University of Puerto Rico at Mayaguez*, 15, 2004. → page 42

[4] T. Adams. Lime kiln principles and operations. *Proceedings of Tappi Kraft Recovery Short Course*, 2:1–18, 1996. → page 65

[5] A. Agrawal and P. Ghoshdastidar. Computer simulation of heat transfer in a rotary lime kiln. *Journal of Thermal Science and Engineering Applications*, 10(3), 2018. → page 81

[6] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019. → page 100

[7] K. Anand, E. Mamatha, C. S. Reddy, and M. Prabha. Design of neural network based expert system for automated lime kiln system. *Journal Européen des Systèmes Automatisés*, 52(4):369–376, 2019. → page 14

[8] F. Azgomi, R. Farnood, and H. Tran. Effect of liming ratio on lime mud settling and filterability. In *Proc. of the Annual Research Review Meeting on Increasing Energy and Chemical Recovery Efficiency in the Kraft Process*, 2008. → page 65

[9] A. Bathelt, N. L. Ricker, and M. Jelali. Revision of the tennessee eastman process model. *IFAC-PapersOnLine*, 48(8):309–314, 2015. → page 16

[10] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013. → pages 17, 19, 45

[11] S. Bennett. A brief history of automatic control. *IEEE Control Systems Magazine*, 16(3):17–25, 1996. → page 26

[12] J. Bergstra, D. Yamins, and D. Cox. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *International conference on machine learning*, pages 115–123, 2013. → page 51

[13] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl. Algorithms for hyper-parameter optimization. In *Advances in neural information processing systems*, pages 2546–2554, 2011. → page 51

[14] BlueMarvel. Heatseeker. https://www.bluemarvel.ai/heatseeker. → pages 25, 70

[15] B. E. Bogdanski. The rise and fall of the canadian pulp and paper sector. *The Forestry Chronicle*, 90(6):785–793, 2014. → page 64

[16] J. Bouvrie. Notes on convolutional neural networks. 2006. → page 20

[17] S. L. Brunton and J. N. Kutz. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2019. → page 48

[18] G. S. Chadha, M. Krishnamoorthy, and A. Schwung. Time series based fault detection in industrial processes using convolutional neural networks. In *IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*, volume 1, pages 173–178. IEEE, 2019. → pages xiv, 19

[19] G. S. Chadha, A. Panambilly, A. Schwung, and S. X. Ding. Bidirectional deep recurrent neural networks for process fault classification. *ISA transactions*, 106:330–342, 2020. → pages xiv, 19

[20] S. Chen, J. Yu, and S. Wang. One-dimensional convolutional auto-encoder-based feature learning for fault diagnosis of multivariate processes. *Journal of Process Control*, 87:54–67, 2020. → pages xiv, 19

[21] F. Cheng, Q. P. He, and J. Zhao. A novel process monitoring approach based on variational recurrent autoencoder. *Computers & Chemical Engineering*, 129:106515, 2019. → pages xiv, 19

[22] L. H. Chiang, E. L. Russell, and R. D. Braatz. *Fault detection and diagnosis in industrial systems*. Springer Science & Business Media, 2000. → page 11

[23] T. T. Chmelyk. An integrated approach to model predictive control of an industrial lime kiln. 2002. → page 94

[24] A. C. C. Coolen. *A Beginner's Guide to the Mathematics of Neural Networks*, pages 13–70. Springer London, London, 1998. → page 18

[25] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20 (3):273–297, 1995. → page 47

[26] J. S. Cramer. The origins of logistic regression. 2002. → page 47

[27] H. Dernegård, H. Brelid, and H. Theliander. Characterization of a dusting lime kiln–a mill study. *Nordic Pulp and Paper Research Journal*, 32(1): 25–34, 2017. → page 65

[28] B. Ding, H. Qian, and J. Zhou. Activation functions and their characteristics in deep neural networks. In *2018 Chinese Control And Decision Conference (CCDC)*, pages 1836–1841. IEEE, 2018. → page 18

[29] J. J. Downs and E. F. Vogel. A plant-wide industrial process control problem. *Computers & chemical engineering*, 17(3):245–255, 1993. → page 16

[30] S. Elnawawi, L. C. Siang, D. L. O'Connor, and R. B. Gopaluni. Interactive visualization for diagnosis of industrial model predictive controllers with steady-state optimizers. *Control Engineering Practice*, 121:105056, 2022. ISSN 0967-0661. → page 111

[31] M. Eriksson, M. Carlborg, and M. Broström. Characterization of ring deposits inside a quicklime producing long rotary kiln. *Energy & Fuels*, 33 (11):11731–11740, 2019. → page 81

[32] M. Fardadi. *Modeling Dust Formation in Lime Kilns*. University of Toronto (Canada), 2010. → page 81

[33] T. Gamer, M. Hoernicke, B. Kloepper, R. Bauer, and A. J. Isaksson. The

autonomous industrial plant-future of process engineering, operations and maintenance. *IFAC-PapersOnLine*, 52(1):454–460, 2019. → page 4

[34] X. Gao, F. Deng, and X. Yue. Data augmentation in fault diagnosis based on the wasserstein generative adversarial network with gradient penalty. *Neurocomputing*, 396:487–494, 2020. → pages xiv, 19

[35] X. Gao, F. Yang, and E. Feng. A process fault diagnosis method using multi-time scale dynamic feature extraction based on convolutional neural network. *The Canadian Journal of Chemical Engineering*, 98(6): 1280–1292, 2020. → pages xiv, 19

[36] P. R. Gareau. *A CFD Study of the Effects of Rings on Flow and Temperature in Lime Kilns*. PhD thesis, University of Toronto (Canada), 2020. → page 81

[37] M. Georgallis. *Mathematical modelling of lime kilns*. PhD thesis, University of British Columbia, 2004. → page 81

[38] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016. → pages 3, 14, 17, 18, 20, 44

[39] J. P. Gorog and W. Leary. Ring removal in rotary kilns used by the pulp and paper industry. *Tappi Journal*, 15(3):205–213, 2016. → pages 65, 80, 81

[40] J. P. Gorog, J. G. Hemrick, H. Walker, W. R. Leary, and M. Ellis. Design of refractory linings for balanced energy efficiency, uptime, and capacity in lime kilns. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2014. → page 68

[41] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou. On the class imbalance problem. In *2008 Fourth international conference on natural computation*, volume 4, pages 192–201. IEEE, 2008. → page 44

[42] M. Gürtürk and H. F. Oztop. Energy and exergy analysis of a rotary kiln used for plaster production. *Applied thermal engineering*, 67(1-2): 554–565, 2014. → page 64

[43] S. Heo and J. H. Lee. Fault detection and classification using artificial neural networks. *IFAC-PapersOnLine*, 51(18):470–475, 2018. → pages xiv, 19, 45

[44] B. Hirtz and D. Marshman. Kiln infrared cameras part 2: Closed loop carbonate control. 2018. → page 95

[45] B. Hirtz, D. Marshman, and C. Sheehan. Kiln infra-red cameras for abnormal situation detection and residual carbonate control. *J-for-Journal of Science & Technology for Forest Products and Processes*, 6(3):28–36, 2017. → pages 66, 69, 95

[46] H. Hotelling. Multivariate quality control. techniques of statistical analysis. *McGraw-Hill, New York*, 1947. → page 12

[47] D. Hurd and J. Kollar. Direct current eaf-a review. *Steel Times*, 219(5):245, 1991. → page 34

[48] D. Hurd and J. Kollar. Data for operating single electrode dc furnaces. 1991. → page 34

[49] A. ISA. Isa-18.2: Management of alarm systems for the process industries. *International Society of Automation. Durham, NC, USA*, 2009. → page 112

[50] A. K. Jain, J. Mao, and K. M. Mohiuddin. Artificial neural networks: A tutorial. *Computer*, 29(3):31–44, 1996. → page 17

[51] L. Jiang, Z. Ge, and Z. Song. Semi-supervised fault classification based on dynamic sparse stacked auto-encoders model. *Chemometrics and Intelligent Laboratory Systems*, 168:72–83, 2017. → pages xiv, 19

[52] X. Jiang and Z. Ge. Data augmentation classifier for imbalanced fault classification. *IEEE Transactions on Automation Science and Engineering*, 2020. → pages xiv, 19

[53] S. Joe Qin. Statistical process monitoring: basics and beyond. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(8-9):480–502, 2003. → page 11

[54] R. T. Jones. Dc arc furnaces—past, present, and future. In *Celebrating the Megascale*, pages 129–139. Springer, 2014. → page 34

[55] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015. → page 1

[56] M. Joswiak, Y. Peng, I. Castillo, and L. H. Chiang. Dimensionality reduction for visualizing industrial chemical process data. *Control Engineering Practice*, 93:104189, 2019. → page 98

[57] L. Kanal. Perceptrons. In N. J. Smelser and P. B. Baltes, editors, *International Encyclopedia of the Social & Behavioral Sciences*, pages

11218 – 11221. Pergamon, Oxford, 2001. ISBN 978-0-08-043076-8. doi:https://doi.org/10.1016/B0-08-043076-7/00572-6. URL http://www.sciencedirect.com/science/article/pii/B0080430767005726. → page 17

[58] M. Kano and M. Ogawa. The state of the art in chemical process control in japan: Good practice and questionnaire survey. *J. Process Control*, 2010. → page 118

[59] E. Keim, J. Zuniga, and H. Tran. Combatting lime kiln ringing problems at the arauco constitucion mill. *TAPPI JOURNAL*, 19(7):345–354, 2020. → page 70

[60] E. Keskinkilic. Nickel laterite smelting processes and some examples of recent possible modifications to the conventional route. *Metals*, 9(9):974, 2019. → page 34

[61] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas. Data preprocessing for supervised leaning. *International Journal of Computer Science*, 1(2): 111–117, 2006. → page 37

[62] I. Kotze. Pilot plant production of ferronickel from nickel oxide ores and dusts in a dc arc furnace. *Minerals Engineering*, 15(11):1017–1022, 2002. → page 34

[63] M. Kovačič, K. Stopar, R. Vertnik, and B. Šarler. Comprehensive electric arc furnace electric energy consumption modeling: A pilot study. *Energies*, 12(11):2142, 2019. → page 7

[64] B. Krawczyk. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, 5(4):221–232, Nov 2016. ISSN 2192-6360. doi:10.1007/s13748-016-0094-0. URL https://doi.org/10.1007/s13748-016-0094-0. → page 44

[65] H. Lagendijk and R. Jones. Production of ferronickel from nickel laterites in a dc arc furnace. Mintek, 1997. → page 34

[66] Y. Le Cun, L. D. Jackel, B. Boser, J. S. Denker, H. P. Graf, I. Guyon, D. Henderson, R. E. Howard, and W. Hubbard. Handwritten digit recognition: Applications of neural network chips and automatic learning. *IEEE Communications Magazine*, 27(11):41–46, 1989. → page 19

[67] Y. Le Cun, O. Matan, B. Boser, J. S. Denker, D. Henderson, R. E. Howard,

W. Hubbard, L. Jacket, and H. S. Baird. Handwritten zip code recognition with multilayer networks. In *[1990] Proceedings. 10th International Conference on Pattern Recognition*, volume 2, pages 35–40. IEEE, 1990. → page 19

[68] L. Le Guen and F. Huchet. Thermal imaging as a tool for process modelling: application to a flight rotary kiln. *Quantitative InfraRed Thermography Journal*, 17(2):79–95, 2020. → page 66

[69] Y. Lecun. A theoretical framework for back-propagation. 08 2001. → page 18

[70] Y. LeCun, Y. Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361 (10):1995, 1995. → page 19

[71] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553): 436–444, 2015. → page 2

[72] G. Lee, T. Tosukhowong, J. H. Lee, and C. Han. Fault diagnosis using the hybrid method of signed digraph and partial least squares with time delay: The pulp mill process. *Industrial & engineering chemistry research*, 45 (26):9061–9074, 2006. → page 13

[73] K. B. Lee, S. Cheon, and C. O. Kim. A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing*, 30(2): 135–142, 2017. → page 49

[74] D. Li, H. Wang, and J. Zhou. Novelty detection for multimode process using gans with learning disentangled representation. In *2020 Chinese Control And Decision Conference (CCDC)*, pages 2536–2541. IEEE, 2020. → pages xiv, 19

[75] Y. Li. Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*, 2017. → page 2

[76] S. C. Lim, S. Elnawawi, L. D. Rippon, D. L. O'Connor, and R. B. Gopaluni. Data quality over quantity: Pitfalls and guidelines for process analytics. *IFAC World Congress 2023*, pages 1–8, 2023. → pages xix, 112

[77] J. Lindblom and H. Theliander. The influence of carbon dioxide on ring

and ball formation in a pilot-scale rotary kiln. *KONA Powder and Particle Journal*, 19:109–117, 2001. → page 79

[78] R. Longadge and S. Dongre. Class imbalance problem in data mining review. *ArXiv*, abs/1305.1707, 2013. → page 44

[79] Q. Lu, R. B. Gopaluni, M. G. Forbes, P. D. Loewen, J. U. Backström, and G. A. Dumont. Model-plant mismatch detection with support vector machines. *IFAC-PapersOnLine*, 50(1):7993–7998, 2017. → page 48

[80] Q. Lu, L. D. Rippon, R. B. Gopaluni, M. G. Forbes, P. D. Loewen, J. Backström, and G. A. Dumont. Noncausal modeling and closed-loop optimal input design for cross-directional processes of paper machines. In *2017 American Control Conference (ACC)*, pages 2837–2842. IEEE, 2017.

[81] F. Lv, C. Wen, Z. Bao, and M. Liu. Fault diagnosis based on deep learning. In *2016 American Control Conference (ACC)*, pages 6851–6856. IEEE, 2016. → pages xiv, 17, 19, 45, 82

[82] F. Lv, C. Wen, M. Liu, and Z. Bao. Weighted time series fault diagnosis based on a stacked sparse autoencoder. *Journal of Chemometrics*, 31(9): e2912, 2017. → pages xiv, 19

[83] M. A. Martins, L. S. Oliveira, and A. S. Franca. Modeling and simulation of limestone calcinations in rotary kilns. *ZKG International*, 55(4):76–87, 2002. → page 81

[84] L. McInnes, J. Healy, and J. Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018. → page 98

[85] W. Meihack. The potential role of fluidized beds in the metallurgical industry. *Journal of the Southern African Institute of Mining and Metallurgy*, 86(5):153–160, 1986. → page 34

[86] G. Millbank and K. McPeek-Newman. *Intercontinental Pulp and Paper Training Manual - Lime Kiln*. PRAXIS Publishing Limited, 1991. → page 65

[87] G. Mohanan. *An Experimental Study on the Effects of Rings on Flame Length and Stability in Lime Kilns*. PhD thesis, University of Toronto (Canada), 2017. → page 81

[88] M. A. Muhammed and A. J. Al-Yasiri. Estimating the thickness of coating

in the burning zone of cement kilns including the aging factor. *The Iraqi Journal For Mechanical And Material Engineering*, 12(3), 2012. → page 81

[89] S. Narasimhan and S. L. Shah. Model identification and error covariance matrix estimation from noisy data using pca. *Control Engineering Practice*, 16(1):146–155, 2008. → page 12

[90] K. Nesselrodt, S. Lee, J. D. Andrews, and P. W. Hart. Mill study on improving lime kiln efficiency. *TAPPI JOURNAL*, 14(2):133–139, 2015. → page 78

[91] K. Nouman, Z. Asim, and K. Qasim. Comprehensive study on performance of pid controller and its applications. In *2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pages 1574–1579. IEEE, 2018. → page 26

[92] NRCan. Benchmarking energy intensity in the canadian steel industry. https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/oee/files/pdf/industrial/SteelBenchmarkEnglish.pdf, 2007. → page 34

[93] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall. Activation functions: Comparison of trends in practice and research for deep learning, 2018. → page 18

[94] N. R. C. O. of Energy Efficiency. Benchmarking energy use in canadian pulp and paper mills in collaboration with the pulp and paper research institute of canada. publications.gc.ca/pub?id=9.667229&sl=0, 2008. → page 63

[95] M. Onel, C. A. Kieslich, Y. A. Guzman, C. A. Floudas, and E. N. Pistikopoulos. Big data approach to batch process monitoring: Simultaneous fault detection and diagnosis using nonlinear support vector machine-based feature selection. *Computers & chemical engineering*, 115: 46–63, 2018. → page 48

[96] K. Pearson. Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 6(2):559, 1901. → page 12

[97] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay.

Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. → pages 52, 101

[98] K. Peltonen. Kiln ring formation– a review of what (we think) we know, 2015. → page 81

[99] S. J. Qin. Process data analytics in the era of big data. *AIChE Journal*, 60 (9):3092–3100, 2014. → page 4

[100] T. W. Rauber, L. H. Mello, V. F. Rocha, and F. M. Varejão. Multi-label fault classification experiments in a chemical process. In *2014 Brazilian Conference on Intelligent Systems*, pages 265–270. IEEE, 2014. → page 47

[101] X. Ren, Y. Zou, and Z. Zhang. Fault detection and classification with feature representation based on deep residual convolutional neural network. *Journal of Chemometrics*, 33(9):e3170, 2019. → pages xiv, 19

[102] L. Rippon. Lime kiln visualization tutorial. https://github.com/LeeRippon/KilnVisual, 2020. → page 79

[103] L. Rippon, B. Hirtz, C. Sheehan, T. Reinheimer, P. Loewen, and B. Gopaluni. Visualization of multiscale ring formation in a rotary kiln. *Nordic Pulp & Paper Research Journal*, 36(4):549–558, 2021. → pages xvi, 25, 66, 67, 111

[104] L. D. Rippon. Detection and diagnosis of ring formation in rotary lime kilns - part i. https://youtu.be/zLHGEO69wZc. June 8, 2021. → pages viii, 63

[105] L. D. Rippon, Q. Lu, M. G. Forbes, R. B. Gopaluni, P. D. Loewen, and J. U. Backström. Machine direction adaptive control on a paper machine. *Industrial & Engineering Chemistry Research*, 58(26):11452–11473, 2019. → pages 48, 89

[106] L. D. Rippon, I. Yousef, R. B. Gopaluni, B. Hosseini, J. F. Beaulieu, C. Prévost, and S. L. Shah. Process analytics and machine learning to predict arc loss in an electric arc furnace. In *59th Conference of Metallurgists 2020 hosting the 4th International Uranium Conference*, 2020. → pages xiv, 34, 35

[107] L. D. Rippon, B. Hirtz, C. Sheehan, T. Reinheimer, C. van der Merwe, P. Loewen, and B. Gopaluni. Detection and diagnosis of ring formation in

rotary lime kilns. *Canadian Chemical Engineering Conference Proceedings*, pages 23–29, 2021. → pages xvi, xvii, 68, 85

[108] L. D. Rippon, I. Yousef, B. Hosseini, A. Bouchoucha, J. F. Beaulieu, C. Prévost, M. Ruel, S. Shah, and R. B. Gopaluni. Representation learning and predictive classification: Application with an electric arc furnace. *Computers & Chemical Engineering*, 150:107304, 2021. → pages xiv, xv, 20, 42, 49, 50

[109] L. D. Rippon, B. Hirtz, C. Sheehan, T. Reinheimer, C. van der Merwe, P. Loewen, and B. Gopaluni. Rotary kiln monitoring with shell temperature visualization and process analytics. In *2022 TAPPI PEERS and IBBC Conference Proceedings*. TAPPI Press, 2022. → pages 66, 89

[110] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, pages 65–386, 1958. → page 17

[111] E. L. Russell, L. H. Chiang, and R. D. Braatz. Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 51 (1):81–93, 2000. → page 12

[112] E. L. Russell, L. H. Chiang, and R. D. Braatz. *Data-driven methods for fault detection and diagnosis in chemical processes*. Springer Science & Business Media, 2012. → pages 13, 14

[113] K. Severson, P. Chaiwatanodom, and R. D. Braatz. Perspectives on process monitoring of industrial systems. *Annual Reviews in Control*, 42:190–200, 2016. → pages 13, 48

[114] J. Shang, M. Chen, H. Ji, D. Zhou, H. Zhang, and M. Li. Dominant trend based logistic regression for fault diagnosis in nonstationary processes. *Control Engineering Practice*, 66:156–168, 2017. → page 47

[115] C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949. → page 90

[116] R. Sharmin, S. L. Shah, and U. Sundararaj. A pca based fault detection scheme for an industrial high pressure polyethylene reactor. *Macromolecular Reaction Engineering*, 2(1):12–30, 2008. → pages 12, 13

[117] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van

Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016. → page 4

[118] S. Skogestad and I. Postlethwaite. *Multivariable feedback control: analysis and design*, volume 2. Wiley New York, 2007. → page 13

[119] G. A. Smook, M. J. Kocurek, et al. *Handbook for pulp & paper technologists*. Canadian Pulp and Paper Association, 1982. → page 65

[120] P. Spyridon and Y. S. Boutalis. Generative adversarial networks for unsupervised fault detection. In *2018 European Control Conference (ECC)*, pages 691–696. IEEE, 2018. → pages xiv, 19

[121] M. Strathern. 'improving ratings': audit in the british university system. *European review*, 5(3):305–321, 1997. → page 60

[122] W. Sun and R. D. Braatz. Opportunities in tensorial data analytics for chemical and biological manufacturing processes. *Computers & Chemical Engineering*, 143:107099, 2020. → page 1

[123] W. Sun, A. R. Paiva, P. Xu, A. Sundaram, and R. D. Braatz. Fault detection and identification using bayesian recurrent neural networks. *Computers & Chemical Engineering*, 141:106991, 2020. → pages xiv, 19, 118

[124] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018. → page 2

[125] N. F. Thornhill, S. L. Shah, B. Huang, and A. Vishnubhotla. Spectral principal component analysis of dynamic process data. *Control Engineering Practice*, 10(8):833–846, 2002. → page 12

[126] H. Tran. Lime kiln chemistry and effects on kiln operations. *Pulp & Paper Centre and Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Canada*, 2007. → pages 7, 63, 65, 79, 81

[127] H. Tran and D. Barham. An overview of ring formation in lime kilns. *Tappi Journal*, 74(1):131–136, 1991. → page 81

[128] H. Tran, X. Mao, and D. Barham. Mechanisms of ring formation in lime kilns. *Journal of pulp and paper science*, 19(4):J167, 1993. → page 81

[129] Y. Tsai, Q. Lu, L. Rippon, S. Lim, A. Tulsyan, and B. Gopaluni. Pattern

and knowledge extraction using process data analytics: A tutorial. *IFAC-PapersOnLine*, 51(18):13–18, 2018.

[130] J. V. Tu. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology*, 49(11):1225 – 1231, 1996. ISSN 0895-4356. doi:https://doi.org/10.1016/S0895-4356(96)00002-9. URL http://www.sciencedirect.com/science/article/pii/S0895435696000029. → page 17

[131] V. N. Vapnik and A. Y. Chervonenkis. Necessary and sufficient conditions for the uniform convergence of means to their expectations. *Theory of Probability & Its Applications*, 26(3):532–553, 1982. → page 47

[132] V. Venkatasubramanian. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE Journal*, 65(2):466–478, 2019. → page 4

[133] V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri. A review of process fault detection and diagnosis: Part ii: Qualitative models and search strategies. *Computers & chemical engineering*, 27(3):313–326, 2003. → page 12

[134] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin. A review of process fault detection and diagnosis: Part iii: Process history based methods. *Computers & chemical engineering*, 27(3):327–346, 2003. → page 12

[135] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri. A review of process fault detection and diagnosis: Part i: Quantitative model-based methods. *Computers & chemical engineering*, 27(3): 293–311, 2003. → page 12

[136] S. Visa and A. Ralescu. Issues in mining imbalanced data sets-a review paper. In *Proceedings of the sixteen midwest artificial intelligence and cognitive science conference*, volume 2005, pages 67–73. sn, 2005. → page 44

[137] N. Wagner and J. M. Rondinelli. Theory-guided machine learning in materials science. *Frontiers in Materials*, 3:28, 2016. → page 45

[138] K. Wagstaff. Machine learning that matters. *arXiv preprint arXiv:1206.4656*, 2012. → page 15

[139] K. Wang, L. Rippon, J. Chen, Z. Song, and R. B. Gopaluni. Data-driven dynamic modeling and online monitoring for multiphase and multimode batch processes with uneven batch durations. *Industrial & Engineering Chemistry Research*, 58(30):13628–13641, 2019.

[140] J. Weston and C. Watkins. Multi-class support vector machines. Technical report, Citeseer, 1998. → page 48

[141] S. Wold, M. Sjöström, and L. Eriksson. Pls-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2): 109–130, 2001. → pages 13, 14

[142] H. Wu and J. Zhao. Deep convolutional neural network model based chemical process fault diagnosis. *Computers & Chemical Engineering*, 115:185–197, 2018. → pages xiv, 19, 20, 45, 48

[143] D. Xie and L. Bai. A hierarchical deep neural network for fault diagnosis on tennessee-eastman process. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 745–748. IEEE, 2015. → page 45

[144] S. Yan and X. Yan. Using labeled autoencoder to supervise neural network combined with k-nearest neighbor for visual industrial process monitoring. *Industrial & Engineering Chemistry Research*, 58(23):9952–9958, 2019. → pages xiv, 19

[145] X. Yang and D. Feng. Generative adversarial network based anomaly detection on the benchmark tennessee eastman process. In *2019 5th International Conference on Control, Automation and Robotics (ICCAR)*, pages 644–648. IEEE, 2019. → pages xiv, 19, 60

[146] Z. Yi, H. Xiao, and J. Song. An alumina rotary kiln monitoring system based on infrared ray scanning. *Measurement*, 46(7):2051–2055, 2013. → page 66

[147] S. Yin, S. X. Ding, A. Haghani, H. Hao, and P. Zhang. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process. *Journal of process control*, 22(9): 1567–1581, 2012. → page 17

[148] I. Yousef. The arc loss challenge. https://github.com/LeeRippon/KilnVisual, 2023. → page 60

[149] I. Yousef, L. D. Rippon, C. Prévost, S. L. Shah, and R. B. Gopaluni. The arc loss challenge: A novel industrial benchmark for process analytics and machine learning. *Journal of Process Control*, 128:103023, 2023. → pages 7, 16

[150] J. Yu, X. Zheng, and S. Wang. A deep autoencoder feature learning method for process pattern recognition. *Journal of Process Control*, 79:1–15, 2019. → pages xiv, 19

[151] W. Yu and C. Zhao. Broad convolutional neural network based industrial process fault diagnosis with incremental learning capability. *IEEE Transactions on Industrial Electronics*, 67(6):5081–5091, 2019. → page 49

[152] J. Yuan and Y. Tian. An intelligent fault diagnosis method using gru neural network towards sequential data in dynamic processes. *Processes*, 7(3): 152, 2019. → pages xiv, 19

[153] X. Zhang, Y. Zou, S. Li, and S. Xu. A weighted auto regressive lstm based approach for chemical processes modeling. *Neurocomputing*, 367:64–74, 2019. → pages xiv, 19

[154] Z. Zhang, T. Jiang, S. Li, and Y. Yang. Automated feature learning for nonlinear process monitoring–an approach using stacked denoising autoencoder and k-nearest neighbor rule. *Journal of Process Control*, 64: 49–61, 2018. → pages xiv, 19

[155] H. Zhao, S. Sun, and B. Jin. Sequential fault diagnosis based on lstm neural network. *IEEE Access*, 6:12929–12939, 2018. → pages xiv, 19, 21

[156] Z. Zheng, X. Wu, and R. Srihari. Feature selection for text categorization on imbalanced data. *ACM Sigkdd Explorations Newsletter*, 6(1):80–89, 2004. → page 44

Appendix A

Supporting Materials

Supporting material related to the pyrometallurgy and kraft pulping case studies is provided in the following two sections.

A.1 Pyrometallurgy Case Study

The operating variables in the pyrometallurgy dataset used to introduce the arc loss fault, develop a predictive inferential sensor, and create the arc loss FDD benchmark challenge are described in Table A.1.

Table A.1: Overview of operating variables analyzed to predict arc loss for the pyrometallurgy case study.

Description	Range	Unit
Smelting parameters		
Electrode A power	0-60	MW
Electrode B power	0-60	MW
Total power	0-100	MW
Electrode A power set point	0-60	MW
Electrode B power set point	0-60	MW

To be continued

Table A.1 (continued)

Description	Range	Unit
Electrode A current	0-100	kA
Electrode B current	0-100	kA
Electrode A current set point	0-100	kA
Electrode B current set point	0-100	kA
Electrode A voltage	$<$2200	Volts
Electrode B voltage	$<$2200	Volts
Electrode A voltage set point	$\leq$2200	Volts
Electrode B voltage set point	$\leq$2200	Volts
Resistance around electrode A	≥ 0	mΩ
Resistance around electrode B	≥ 0	mΩ
Resistance around electrode A set point	≥ 0	mΩ
Resistance around electrode B set point	≥ 0	mΩ
Specific Energy Ratio	410-440	W/ton
Arc A length		mm
Arc B length		mm
Crucible (the wall) heat loss	0-5	MW
Roof heat loss	≥ 0	MW
Plain cooler heat loss	≥ 0	MW
Upper chilled water heat loss	≥ 0	MW
Lower chilled water heat loss	≥ 0	MW
Hearth fans heat loss	≥ 0	MW
Heat loss through tap chutes in hearth	≥ 0	MW
Slag level		mm
Metal level		mm
Off-gas temperature	180-630	$^{\circ}$C
Slag tap A valve opening	0-100	%
Slag tap B valve opening	0-100	%
CO_2 volume	0-25	%

To be continued

Table A.1 (continued)

Description	Range	Unit
Slag temperature after being tapped		°C
Furnace feed parameters		
Furnace feed rate	0-200	TPH
Furnace feed inlet diverter 5A position	[0,1]	
Furnace feed inlet diverter 5B position	[0,1]	
Furnace feed inlet diverter 5C position	[0,1]	
Furnace feed inlet diverter 5D position	[0,1]	
Furnace feed inlet diverter 5E position	[0,1]	
Furnace feed inlet diverter 5F position	[0,1]	
Weir 1 valve opening	[0,1]	
Weir 2 valve opening	[0,1]	
Weir 3 valve opening	[0,1]	
Weir 4 valve opening	[0,1]	
Weir 5 valve opening	[0,1]	
Weir 6 valve opening	[0,1]	
Weir 7 valve opening	[0,1]	
Weir 8 valve opening	[0,1]	
Weir 1 flow rate	0-200	TPH
Weir 2 flow rate	0-200	TPH
Weir 3 flow rate	0-200	TPH
Weir 6 flow rate	0-200	TPH
Weir 7 flow rate	0-200	TPH
Weir 8 flow rate	0-200	TPH
Weir 1 flow rate	≥ 0	mA
Weir 2 flow rate	≥ 0	mA
Weir 3 flow rate	≥ 0	mA
Weir 4 flow rate	≥ 0	mA
Weir 5 flow rate	≥ 0	mA

To be continued

Table A.1 (continued)

Description	Range	Unit
Weir 6 flow rate	≥ 0	mA
Weir 7 flow rate	≥ 0	mA
Weir 8 flow rate	≥ 0	mA
Total microwave flow rate	≥ 0	mA
Feed temperature after leaving weir 1		°C
Feed temperature after leaving weir 2		°C
Feed temperature after leaving weir 3		°C
Feed temperature after leaving weir 4		°C
Feed temperature after leaving weir 5		°C
Feed temperature after leaving weir 6		°C
Feed temperature after leaving weir 7		°C
Feed temperature after leaving weir 8		°C
Feed temperature in distribution bin 1		°C
Feed temperature in distribution bin 2		°C
Feed temperature in distribution bin 3		°C
Feed temperature in distribution bin 4		°C
Feed temperature entering port A		°C
Feed temperature entering port B1		°C
Feed temperature entering port B2		°C
Feed temperature entering port B3		°C
Feed temperature entering port B4		°C
Feed temperature entering port B5		°C
Feed temperature entering port B6		°C
Feed temperature entering port C1		°C
Feed temperature entering port C2		°C
Feed temperature entering port C3		°C
Feed temperature entering port C4		°C
Feed temperature entering port C5		°C
Feed temperature entering port C6		°C

To be continued

Table A.1 (continued)

Description	Range	Unit
Feed temperature entering port C7		°C
Feed temperature entering port C8		°C
Feed temperature entering port C9		°C
Feed temperature entering port C10		°C
Weigh bin cone valve position feedback	≥ 0	mA
Weigh bin cone valve opening measured value	0-100	%
Weigh bin cone valve position controller output	≥ 0	mA
Cone valve position controller set point	0-100	%
Furnace feed pipe A, A-port flow	≥ 0	TPH
Reduction parameters		
Fluidized bed reducer level set point		m
Fluidized bed reducer level		m
Fluidized bed reducer level		mA
Fluidized bed reducer level controller output	≥ 0	mA
Fluidized bed reducer cone valve control	0-100	%
Fluidized bed reducer temperature	800-1100	°C
Calcining parameters		
Calciner feed rate	≥ 0	TPH
Coal feed rate	≥ 0	TPH
Laboratory parameters		
Al_2O_3 concentration in the slag		ppm
FeO concentration in the slag		ppm
MgO concentration in the slag		ppm
Ni concentration in the slag		ppm
SO_2 concentration in the slag		ppm

The relatively small search space over which hyper-parameters are optimized for the traditional ML algorithms is presented in Table A.2.

Table A.2: Hyperparameter search space for traditional ML algorithms.

	LR	L-SVC	K-SVC
λ	(0.001, 0.01, 1, 10, 100)		
tolerance	(0.001, 0.0001, 0.00001)		
penalty	(ℓ_1, ℓ_2, elast.)	(ℓ_1, ℓ_2)	
loss		(hinge, SH)	
kernel			(RBF, poly., sig.)
degree			(2, 3, 4, 5)

The vast search space over which hyper-parameters are optimized for the deep learning algorithms is presented in Table A.3.

Table A.3: Hyperparameter search space for deep learning algorithms.

	ANN	CNN
optimizer	(RMSprop, Adagrad, Adam, Adadelta, Adamax, SGD)	
λ	(0.1, 0.01, 0.001, 0.0001)	
FCL activation	(relu, tanh, selu, elu)	
no. of FCLs	(1, 2, …, 10)	(1, 2, 3, 4)
FCL size	(32, 64, 128, 256, 512)	(32, 64, 128)
batch size	(32, 64, 128)	(16, 32, 64, 128)
epochs	(25, 35)	(20, 30, 40)
no. of CLs		(1, 2)
CL activation		(relu, tanh, selu, elu)
filters		(8, 16, 32, 64)
filter size		[(3,3), (5,5)]
pool size		[(2,2), (4,4)]

The precision (i.e., PPV), F_1 score and F_β ($\beta = 0.25$) score for each experimental configuration are provided as supplementary result metrics in Table A.4.

Table A.4: Supplemental experimental result metrics.

		TP	FP	TN	FN	PPV	F_1	F_β
NA	LR	182	49	151	76	0.788	0.744	**0.708**
	L-SVC	185	46	141	86	0.801	0.737	0.687
	K-SVC	167	64	156	71	0.723	0.712	0.702
	ANN	176	55	152	75	0.762	0.730	0.703
	CNN	181	50	141	86	0.784	0.727	0.681
PCA $d = 41$	LR	186	45	148	79	**0.805**	**0.750**	0.705
	L-SVC	176	55	151	76	0.762	0.729	0.701
	K-SVC	144	87	158	69	0.623	0.649	0.674
	ANN	176	55	142	85	0.762	0.715	0.677
	CNN	181	50	149	78	0.784	0.739	0.702
PLS $d = 16$	LR	184	47	130	97	0.797	0.719	0.659
	L-SVC	169	62	146	81	0.732	0.703	0.678
	K-SVC	166	65	149	78	0.719	0.699	0.682
	ANN	166	65	147	80	0.719	0.696	0.676
	CNN	147	84	161	66	0.636	0.662	0.688

A.2 Kraft Pulping Case Study

The operating variables in the kraft pulping dataset used to investigate ring formation and develop a residual calcium carbonate inferential sensor are described in Table A.5.

Table A.5: Overview of operating variables analyzed to investigate ring formation and predict residual calcium carbonate for the kraft pulping case study.

Description	Unit
Lime mud filtering parameters	
Filter 1 mass percent solids	%
Filter 1 mud flow rate	TPD
Filter 1 vacuum pressure	in. WC
Filter 2 mass percent solids	%
Filter 2 mud flow rate	TPD
Filter 2 vacuum pressure	in. WC
Kiln mud feed target	TPD
Kiln mud feed rate	TPD
Scrubber mud slurry recycle flow 1	L/s
Scrubber mud slurry recycle flow 2	L/s
Mud slurry recycle density	kg/L
Mud slurry recycle mass percent solids	%
Kiln operating parameters	
Kiln feed end temperature 1	°C
Kiln feed end temperature 2	°C
Kiln feed end pressure	in. WC
ID fan inlet temperature	°C
ID fan speed	RPM
Lime kiln speed	RPM
Kiln drive amps	Amps
Kiln feed end oxygen 1	%

To be continued

Table A.5 (continued)

Description	Unit
Kiln feed end oxygen 2	%
Mid-zone temperature 1	°C
Mid-zone temperature 2	°C
Natural gas flow to burner	m^3/h
Primary air to kiln	m^3/h
Kiln firing end temperature 1	°C
Kiln firing end temperature 2	°C
Kiln firing end pressure	in. WC
Residual calcium carbonate	%
Thermal IR *camera data*	
Kiln bearing shield	°C
Kiln beam	°C
Kiln gearbox	°C
Kiln dump gate	°C
Shell hot spot 0-90ft	°C
Shell hot spot 0-180ft	°C
Kiln shell temperature 15 ft	°C
Kiln shell temperature 20 ft	°C
Kiln shell temperature 30 ft	°C
Kiln shell temperature 37 ft	°C
Kiln shell temperature 43 ft	°C
Kiln shell temperature 50 ft	°C
Kiln shell temperature 60 ft	°C
Kiln shell temperature 80 ft	°C
Kiln shell temperature 100 ft	°C
Kiln shell temperature 113 ft	°C
Kiln shell temperature 120 ft	°C
Kiln shell temperature 133 ft	°C

To be continued

Table A.5 (continued)

Description	Unit
Kiln shell temperature 150 ft	°C
Kiln shell temperature 154 ft	°C
Kiln shell temperature 157 ft	°C
Kiln shell temperature 160 ft	°C
Kiln shell temperature 195 ft	°C
Kiln shell temperature 210 ft	°C
Kiln shell temperature 220 ft	°C
Kiln shell temperature 230 ft	°C
Kiln shell temperature 240 ft	°C
Kiln shell temperature 249 ft	°C
Kiln shell temperature 254 ft	°C
Kiln shell temperature 259 ft	°C
Recausticizing parameters	
White liquor sulphidity 1	%
White liquor sulphidity 2	%
Causticizer 1 settling time	#
White liquor causticizing efficiency	%
Clarifier 1 causticizing efficiency	%
Kiln mud feed target	TPD
Kiln mud feed rate	TPD
Weak wash total titratable alkali	g/L
Ambient conditions	
Wind speed	ft/min
Ambient temperature	°C

The majority of the hyperparameter search space for the feature learning and regression models implemented in the hyper-inferential sensing case study are presented in Table A.6 and Table A.7. Table A.6 provides hyperparameters for traditional (or non deep learning) algorithms and Table A.7 provides the hyperpa-

rameters for the deep learning algorithms which share common search spaces for epochs, learning rate, and batch size.

Table A.6: Hyperparameter search space for traditional ML algorithms.

	PCA	PLS	*k*NN	XGBoost
Components	(10,...,500)	(2,...,input size−1)		
Neighbors			(1,...,20)	
Estimators				(10,...,50/200)
Max depth				(3,...,10/20)
Learning rate				(0.0001,...,1)

Table A.7: Hyperparameter search space for deep learning algorithms.

	AE	ANN	CNN	LSTM
Epochs	(10, ..., 100)			
Learning rate	(0.0001, ..., 0.01)			
Batch size	(16, ..., 128)			
no. hidden layers	(1,1)	(1,...,8)	(2)	(1,...,8)
Hidden size(s)	(32,...,256)	(16,...,128)		(32,...,128)
Encoding size	(8, ..., hidden size)			
CNN filters 1			(8, 32)	
CNN filters 2			(16, 64)	