EMPIRICAL STUDY OF THE EFFECT OF STOCHASTIC VARIABILITY ON THE PERFORMANCE OF HUMAN-DEPENDENT FLEXIBLE FLOW LINES

Adam Aboutaleb

A Thesis Submitted in Partial Fulfilment of the Requirement of De Montfort University for the Degree of Doctor of Philosophy

October 2015

ABSTRACT

Manufacturing systems have developed both physically and technologically, allowing production of innovative new products in a shorter lead time, to meet the 21st century market demand. Flexible flow lines for instance use flexible entities to generate multiple product variants using the same routing. However, the variability within the flow line is asynchronous and stochastic, causing disruptions to the throughput rate.

Current autonomous variability control approaches decentralise the autonomous decision allowing quick response in a dynamic environment. However, they have limitations, e.g., uncertainty that the decision is globally optimal and applicability to limited decisions.

This research presents a novel formula-based autonomous control method centered on an empirical study of the effect of stochastic variability on the performance of flexible human-dependent serial flow lines. At the process level, normal distribution was used and generic nonlinear terms were then derived to represent the asynchronous variability at the flow line level. These terms were shortlisted based on their impact on the throughput rate and used to develop the formula using data mining techniques.

The developed standalone formulas for the throughput rate of synchronous and asynchronous human-dependent flow lines gave steady and accurate results, higher than closest rivals, across a wide range of test data sets. Validation with continuous data from a real-world case study gave a mean absolute percentage error of 5%.

The formula-based autonomous control method quantifies the impact of changes in decision variables, e.g., routing, arrival rate, etc., on the global delivery performance target, i.e., throughput, and recommends the optimal decisions independent of the performance measures of the current state. This approach gives robust decisions using pre-identified relationships and targets a wider range of decision variables.

The performance of the developed autonomous control method was successfully validated for process, routing and product decisions using a standard 3x3 flexible flow line model and the real-world case study. The method was able to consistently reach the optimal decisions that improve local and global performance targets, i.e., throughput, queues and utilisation efficiency, for static and dynamic situations. For the case of parallel processing which the formula cannot handle, a hybrid autonomous control method, integrating the formula-based and an existing autonomous control method, i.e., QLE, was developed and validated.

ACKNOWLEDGEMENTS

First, I thank God for giving me the strength and inspirations throughout this three-year journey.

My deepest appreciation is extended to my family for their endless support and love.

I take this opportunity to express my immense gratitude to my PhD supervisors; Riham Khalil for her invaluable guidance and sincere faith, support and encouragement; Dave Stockton for proposing the main idea of this project; Parminder Kang for the fruitful technical discussions and support and Raouf Hamzaoui for his guidance and sharing his expertise.

I would like to thank the InnovateUK (Grant Reference NO: 18834-132285) and De Montfort University for funding this research and Costain for their support of information and data to this research.

Finally, I extend my warm thanks to the great people at the Lean Engineering Research Group. You have a great potential for success – keep it up! Special thanks go to Abdulla Al-Dhaheri, Abdulhamid Askri and Yamini Raju. Wish you all the best of luck in your life and career. You will be missed!

DEDICATION

To my parents and my niece, Anne

DECLARATION

I hereby declare that the work described within this thesis was originally undertaken by me, Adam Aboutaleb, between the dates of registration for the degree of Doctor of Philosophy at De Montfort University, October 2012 to October 2015.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANOVA	Analysis of Variance
CI	Confidence Interval
CRB	Central Reservation Barrier
DES	Discrete Event Simulation
DLRP	Distributed Logistics Routing Protocol
DOF	Degree of Freedom
GA	Genetic Algorithm
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MAXAE	Maximum Mean Absolute Error
MAXAPE	Maximum Mean Absolute Percentage Error
MINAE	Minimum Mean Absolute Error
MINAPE	Minimum Mean Absolute Percentage Error
MLE	Maximum Likelihood Estimator
MTO	Make to Order
MTS	Make to Stock
MTTF	Mean Time to Failure
MTTR	Mean Time to Repair
NN	Artificial Neural Network
OLS	Ordinary Least Squares
PEB	Past Events Based
pdf	Probability density function
QLE	Queue Length Estimator
TPS	Toyota Production System

LIST OF SYMBOLS

N	Number of processes
a	Counter of data set within test set
$\boldsymbol{\beta}$	Regression coefficients
c	Coefficient of variation/coefficient of variation of mean processing times for synchronous flow lines
c_a	Coefficient of variation of arrival rates
c_{av}	Average coefficient of variation
c_b	Coefficient of variation of batch times of concrete
$cScore$	Score of the errors in c
$\mathbf{D}_x$	Data set
e	MAPE
$\mathbf{E}$	MAPE matrix
ε	Regression error term
η_u	Utilisation efficiency
h	sub-set number
i	Location of process
j	Counter of supporting predictors in training set
m	Machine learning method counter
l	Location ratio of the process with maximum mean processing time
λ	Failure rate
μ	Mean processing time
μ_a	Mean arrival rate
μ_b	Mean batch time of concrete
μ_c	Mean concrete load conditioning time
μ_{del}	Mean truck delivery time
μ_{dis}	Mean concrete load discharge time
μ_e	Mean of the %errors of each method and data set within the test set

$\mu_{\max}$	Maximum mean processing time
$\mu_{\min}$	Minimum mean processing time
$\mu Score$	Score of the errors in μ
n	Counter of data sets in training set
$v_{\mu_{del}}$	Data variations of mean truck delivery times
$v_{\sigma_{del}}$	Data variations of standard deviation of truck delivery times
P	Process
$\mathbf{P}$	Predictors set
p	Predictor
q	Number of variability scenarios
q_{T_a}	Number of variability scenarios within test set
Q	Queue Size
QT	Queue time
$\mathbf{R}_w$	Training experiment
$\mathbf{S}$	Train set
σ	Standard deviation
σ_c	Standard deviation of concrete load conditioning times
σ_{del}	Standard deviation of truck delivery times
σ_{dis}	Standard deviation of concrete load discharge times
σ_e	Standard deviation of the %errors of each method and data set within the test set
ss	Sub-set
t	Time
T	Throughput
$\mathbf{T}$	Test set
$\mathbf{T}_o$	Main test set
$\mathbf{T}_u$	Supporting test set
TR	Throughput rate
$\mathbf{TR}^{asyn}$	Throughput rate of asynchronous flow line
TR^d	Throughput rate of deterministic flow line

TR^{e}	Throughput rate of exponential flow line
TR^{in}	Throughput rate of initial value of variability factor
TR^{ideal}	Ideal throughput rate
$\mathbf{TR}^{norm}$	Normalised throughput rate
$\mathbf{TR}^{pred}$	Predicted throughput rate
$\mathbf{TR}^{sim}$	Simulated throughput rate
$\mathbf{TR}^{\mathbf{sync}}$	Throughput rate of synchronous flow line
TT	Throughput time
u	Utilisation
$\mathbf{V}$	Regression covariates
w	Training set number
x	Data set counter
y	Predictor counter

1 INTRODUCTION

1.1 Motivation

During the past decades, several manufacturing systems were developed to keep pace with the significant advancements in technology and telecommunications and tailor products and services to achieve the main strategic goal '*to satisfy the customer requirements*' (Upton 1994). Customer requirements tend to be trending upwards in terms of complexity, which requires reshaping the manufacturing process to be flexible enough to handle a variety of complex products (He et al. 2014).

Manufacturing systems developed significantly over the past years to become more lean, customer-oriented and intelligent (Jasti and Kodali 2015). To cope with the fast-track market changes, flexibility has been built in every element of the manufacturing system (Jain et al. 2013). Flexible flow lines use multi-skilled process owners or flexible machines for human-dependent and machine-based flow lines respectively (Quadt and Kuhn 2005). Flexible flow lines are a cost-effective solution that integrates the benefits of both mass production and mass customisation strategies (Ambani 2011, Sankar et al. 1997). Such flow lines standardise the serial routing for all product variants while allowing manufacturing flexibility to take place at the process level (Quadt and Kuhn 2007). The flexible flow line can be synchronous if the variability across all processes is common and, more generally, asynchronous when inter-process variability exists (Li and Meerkov 2009).

As a result, variability has increased, generating complexity in process and production planning. With the increased complexity of manufacturing systems, production and process planning to maintain the performance targets becomes a challenging task (Daniel and Guide 2000). Several autonomous control techniques have been developed to cope with performance fluctuations (Windt et al. 2010, Grundstein et al. 2015). Deployment of autonomous control into the flexible flow line creates a real-time representation of the current system state and decides the next step autonomously without human intervention. However, to reach to the right decision, the autonomous decision should evaluate the effect of changes in variability based on this decision on the performance of the flow line. The evaluation process should also be carried out in a time-efficient manner before a new status of the actual flow line or customer orders takes place (Stelson et al. 1996).

Evaluative modelling can give this competitive advantage by linking the variability to the performance targets of interest. Several mathematical, simulation and empirical models (Carrascosa 1995, He et al. 2007, Li et al. 2009, Papadopoulous et al. 2009, Li et al. 2013, Tan et al. 2015) were developed for different types of flow lines. However, for stochastic non-Markovian processes that follow a distribution other than the exponential or phase-type distribution, analytical methods do not exist (Meerkov and Yan 2014) and simulation and empirical approaches were the favourable solutions.

While simulation is usually case-based, closed-form empirical formulas can be generic, simple, time efficient and relationships are easily understood (Blumenfeld 1990, Papadopoulos et al. 2009, Li and Meerkov 2009, Hopp and Spearman 2011, Wang et al. 2014). Empirical formulas can also provide accurate estimations of the throughput rate for the process and production planners to use. Empirical formulas can also be integrated into the autonomous control system to assist in the evaluation process of autonomous decisions and allow for an improved resource efficiency and increased throughput rate.

1.2 Problem Statement and Research Gap

A flow line is a type of manufacturing system with a combination of processes and queues, where physical, e.g., raw materials, or virtual, e.g., orders, parts flow through in a standard routing to be transformed into a final product (Buzacott 2013). The process can be dependent on either machines or people or both. If the machine is the core driver, the process will be borne to interruptions such as setups, machine breakdowns and preventive maintenance. The process variability in this case can be medium to high, depending on the frequency and length of the interruptions (Hopp and Spearman 2011) which are usually unpredictable and assumed exponential (Li and Meerkov 2009). On the other hand, the processing time will tend to be more deterministic, especially in fully automated machines (Li et al. 2013). Flexible human-dependent processes have the advantage of producing a range of products while being less interrupted by setups and breakdowns at the cost of having more stochastic non-exponential processing times (Wang et al. 2014) due to flexibility of the human brain, cognitive functions, skills and emotions (OECD 2007).

This natural variability combined with the one due to product complexity and variety produce normally distributed process variability (Hopp and Spearman 2011) with a coefficient of variation that is typically less than 1.

A simple serial flow line of N processes $P_i, i = 1,2,...,N$ and queues $Q_i, i = 1,2,...,N$ is presented in Figure 1.1. Variability of the throughput rate in the flow line is the combination of intra-variability of each individual process P_i, due to the natural variability and product variety and inter-variability between one process and another due to the product complexity.

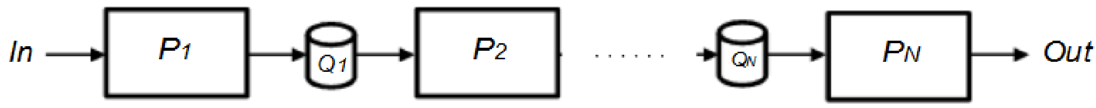


Figure 1.1: Serial Flow Line

The idea of prediction of the impact of variability on the performance of the manufacturing system is a major concern for building autonomous control systems. Decisions regarding process and production planning, e.g., the following processing step to take for a product, depend on the first place on evaluation of the current situation. Although decentralised autonomous control takes decision based on local information, understanding the impact of an autonomous decision on the system-level performance targets can increase the certainty that the decision is optimal (Scholz-Reiter et al. 2009a).

Research in the area of evaluative modelling focused on machine-based flow lines which are widely used in the manufacturing industry. Such models were developed primarily to include queue capacity and repair and failure rates, however, only assume deterministic, exponential or phase based distribution of the processing times (Li et al. 2013), which is not applicable for human-dependent processes such as in the construction industry. Simulation and empirical methods were widely used for the analysis of stochastic non-exponential flow lines, where processes are non-Markovian. Closed-form empirical formulas are usually simple to understand and apply and they can have the potential to model complex flow lines, under few assumptions, without compromising the accuracy of approximation (Papadopoulos 1996).

Empirical formulas were developed (Barten 1962, Anderson and Moodie 1969, Knott 1970, Buxey et al. 1973, Slack and Wild 1980, Muth 1987, Blumenfeld 1990, Medhi 1991, Papadopoulos 1996, Khalil 2005, Blumenfeld and Li 2005, Li and Meerkov 2009 and Hopp and Spearman 2011) for different performance measures, generally system loss and throughput rate of exponential and non-exponential synchronous and asynchronous flow lines. However, current empirical formulas for the throughput rate of non-exponential flow lines (Muth 1987, Blumenfeld 1990, Li and Meerkov 2009) were developed primarily for the synchronous case, i.e., no inter-variability in processing times across the flow line. An exception is an empirical formula developed by Li and Meerkov (2009) for asynchronous non-exponential flow lines, but the formula has a variable TR^e that represents the throughput rate of the exponential flow line, which still needs to be obtained using simulation so the formula cannot be applied on its own to non-exponential flow lines. Wang et al. (2014) and Kang et al. (2015) presented an interesting Markov chain-based analytical model to obtain TR^e for short service-based flow lines with non-exponential processing times. This research opts for the benchmark, i.e., simulation, to determine TR^e and Li and Meerkov (2009) empirical formula was used for comparison purposes.

The empirical work by Li and Meerkov (2005), Li and Meerkov (2009), Meerkov and Yan (2014) shows that the key parameters that play a part to differentiate between exponential and non-exponential flow lines are the maximum mean processing time and the coefficient of variation. This suggests that first principles modelling, based on the parameters of each individual process, is not the best solution and an empirical study is more appropriate. To the best of the author's knowledge, no standalone closed-form empirical formula exists for the throughput rate of asynchronous flow lines with normally distributed process variability.

This research presents an empirical study to determine a standalone closed-form formula of the throughput rate for human-dependent serial flow lines. The closed-form formula is then used as a building block for the control mechanism of an autonomous control system of flexible flow lines.

1.3 Purpose of the Study

The primary concern of this research is to assist production and process planners and engineers to understand and control the effect of process and production planning decisions on the intra- and inter-process variability and hence the system-level performance targets for flexible human-dependent serial flow lines.

This quantitative research is based on an empirical study of the effect of process variability of the throughput rate of synchronous and asynchronous human-dependent serial flow lines. The study exploits data mining and simulation modelling. Synthetic data were used to develop and test the model and actual data, from a case study in construction industry, were used for validation purposes.

1.4 Research Questions

The research problem investigated here is an accurate evaluative model of flexible human-dependent serial flow lines that can be used to autonomously control the flow line based on variability.

This problem raises the following research questions in the light of the existing state-of-the-art:

i. Which data pre-processing method performs the best in handling the bias caused by the simulation error and increases the reliability and confidence in the simulated throughput rate?

ii. How the variability within a non-exponential serial flow line can be represented in a data mining-compatible generic form applicable to asynchronous flow lines with arbitrary length and scenario of intra- and inter-process variability?

iii. Can data mining models produce a simple closed-form formula to estimate the throughput rate of synchronous and asynchronous human-dependent serial flow lines? How accurate will this evaluative model perform for the real-world case study?

iv. Can this formula-based evaluative model be utilised to control the variability within a flexible flow line? If yes, how does it compare to other existing methods in terms of performance and how can it be implemented in a real-world setup?

1.5 Scope and Objectives of the Research

1.5.1 Scope

The scope of this empirical research is to develop an autonomous-decision-support closed-form formula through empirical evaluative models that can help to easily and quickly estimate the effect of each autonomous decision, based on the stochastic variability in process and production planning, on the system-level performance target and use this relationship to control flexible human-dependent serial flow lines.

This research is concerned with closed-form formula-based empirical evaluative modelling of the steady state performance of serial flow lines with reliable machines (processes) and infinite queues using data mining techniques. Figure A.1 of Appendix A (P. A-2) illustrates the scope of the research in terms of evaluative modelling.

1.5.2 Assumptions

The research focuses on flexible flow lines with a standard serial flow line arrangement, infinite queues, if exist, and stochastic processes that follow the rules of normal distribution. This flow line representation primarily agrees with the industrial real-world case study of this research. However, occasionally parallel processing and closed-loops might take place in a real-world setup but they are not the main concern of this research. Furthermore, the research investigates a new method in the autonomous control and validates it with existing techniques. Hence, the research is not concerned at this stage with the actual integration of the research outcomes into current production planning systems.

Hence, the following assumptions are made:

i. The flow lines consist of N serial processes.

ii. The processing time for each process P_i is independent of the upstream and downstream processes P_{i-1} and P_{i+1}, i.e., the flow line is asynchronous.

iii. The time for each process P_i is normally distributed with a mean processing time of $\mu_i, i = 1,2,...,N$ and standard deviation of $\sigma_i, i = 1,2,...,N$.

iv. The human-dependent process P_i is reliable with a failure rate $\lambda_i, i = 1,2,...,N$.

v. Blocking of a process P_i can only occur when it completed processing a part while the downstream process P_{i+1} is still busy and no queue exists between them.

vi. A process P_i can get '*starved*' when the upstream process P_{i-1} is not completed.

vii. Required resources, i.e., machine, people, tools, etc. are always available at the respective process P_i.

viii. If the process P_i is not '*blocked*' or '*starved*', it is in '*busy*' state, i.e., the process P_i is not allowed to be '*idle*'.

ix. The flow line is saturated, i.e., the first process P_1 is never '*starved*' for inputs, e.g., materials, orders, and the last process P_N is never '*blocked*', i.e., it has infinite capacity of inventory.

x. The travel time between processes is zero, i.e., transportation of materials and work in progress is modelled as a separate process.

xi. The loss rate in throughput rate *TR* is zero, i.e., there are no defective products.

1.5.3 Objectives

The research scope is realised through the following main objectives:

i. Generate synthetic data for generic representation of the intra- and inter-process variability within synchronous and asynchronous non-exponential serial flow lines and use Discrete Event Simulation modelling to obtain the steady state simulated throughput rate with high certainty.

ii. Build a Data Mining Framework and use it to develop an empirical formula and perform goodness-of-fit analysis for the estimated throughput rate for synchronous and asynchronous human-dependent serial flow lines.

iii. Build an Autonomous Control Framework for flexible flow lines based on the developed empirical formula.

iv. Validate the developed empirical formula and autonomous control method using representative variability scenarios of flexible flow lines and a real-world case study in the construction industry.

1.6 Structure of the Report

Chapter 2 and 3 cover the literature survey: Chapter 2 starts with an overview of the manufacturing systems, their main milestones of evolution and the challenges they created along the way. The chapter then demonstrates the characteristic advantages of flexible flow lines over the other manufacturing systems and where they fit in the evolution hierarchy. It also explains why evaluative modelling can improve the performance of flexible flow lines to deal with the trending challenges in manufacturing systems. Finally, it gives an overview of the current state-o*f*-the-art in evaluative modelling for flow lines. **Chapter 3** gives more details about the control criteria of autonomous manufacturing systems and how they link to the variability parameters in process and production planning and the performance measures of the system. The chapter also covers the current development within these areas.

Chapter 4 and 5 describe the methodology and methods of this research: Chapter 4 starts with the methodology which will be used during the research. It then covers the two main methodological frameworks of this research, namely Data Mining and Autonomous Control Frameworks, and the methods and steps in each. **Chapter 5** then covers the specific case study used to validate the research outcomes and the methods and steps involved in this case study.

Chapter 6 and 7 are devoted to the results and analysis: Chapter 6 gives the results of the implementation of Data Mining Framework and the outcomes of each step, more importantly the empirical formula for the throughput rate of synchronous and asynchronous serial flow lines with normally distributed process variability. It concludes with the validation of the empirical formula using the real-world case study. **Chapter 7** covers the integration of the developed empirical formula into an autonomous control method and compares the performance of developed autonomous control methods with existing heuristic optimisation and autonomous control methods for flexible flow lines. The chapter concludes with the validation of formula-based autonomous control method within the real-world case study.

Chapter 8 gives the critical evaluation of the research outcomes. It covers the achievements and limitations of this research in pursue of answering the research questions outlined in Section 1.4. Each research outcome is examined in terms of contributions to knowledge over the current state-o*f*-the-art, the precision and thoroughness of results and limitations.

Chapter 9 concludes the research and lists the main points that can be investigated further in light of the outcomes and contributions of this research.

2 EVALUATIVE MODELLING OF FLEXIBLE FLOW LINES

2.1 Introduction

Manufacturing system is the umbrella that includes all facilities and operations within a manufacturing plant. Flow line is the segment of the manufacturing system that deals with the transformation of raw materials to finished products (Oztemel 2010).

During the past few decades, several manufacturing systems have been developed to keep pace with the significant improvements in technology and the major shifts in customer behaviour (Zhang 2009).

A main challenge that originated as a result of the technological advancement and the changes in customer behaviour is the increased variability of the manufacturing system. Hopp and Spearman (2011) classified variability into good and bad variability. The ability of the manufacturing system to produce innovative products on short terms is the good variability. The bad variability is the bi-product of the good one represented by the increased variations within each stage of the manufacturing system.

To analyse the effect of variability based on product complexity and time on the performance targets, researchers have investigated modelling the relationship between variability and performance targets using evaluative models (Papadopoulos 2009). Several mathematical, simulation and empirical methods were developed over the last 60 years.

This chapter gives an overview of the manufacturing systems and their development milestones. It then gives details on how variability affects the manufacturing system. The chapter then covers the state-o*f*-the-art in terms of evaluative modelling of flow lines.

2.2 Overview of Manufacturing Systems

Oztemel (2010) defined manufacturing systems as the "*integration of manufacturing functions such as design, process planning, production planning, quality assurance, storing and shipment, etc.*".

Manufacturing system includes all operations from the design stage of a product to the shipment of the final product. It begins with orders received from and ends with delivery to the customer (Heilala 1999).

Manufacturing system can be divided into the following stages (Oztemel 2010):

i. design;
ii. process planning;
iii. production planning;
iv. manufacturing;
v. quality control; and
vi. storage and shipping.

Figure 2.1 gives the inputs and outputs of each stage of the manufacturing system. The process and production planning are the main concern of this research.

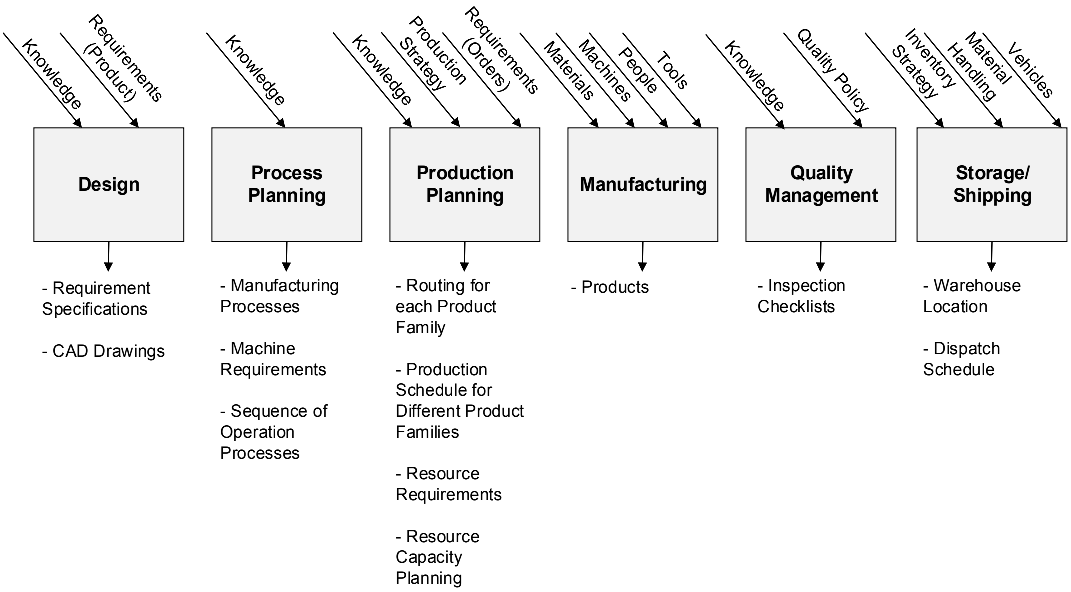


Figure 2.1: Stages of Manufacturing System

2.2.1 Process Planning

Process planning is the translation of design drawings into manufacturing processes using the available resources (Groover 2010).

Process planner will set the processes to be undertaken and their routing to produce each product in the customer's order. An alternative routing for products should also be provided to address the issues due to unforeseen circumstances, e.g., machine breakdown (Chryssolouris 2006). The process planner starts with a routing sheet and uses it to build an operations list for a particular product (Scallan 2003).

The operations list document defines the following parameters (Singh 1995):

i. sequence of processes;
ii. machine or process where product is processed;
iii. machine setup procedure and requirements;
iv. tools to be used; and
v. setup and processing times.

The second to fourth parameters are usually described using the organisation standard terms while the first and last are numeric. In essence, both sequence and setup and processing times represent the variability of manufacturing process, as changes to other parameters will lead back to changes to these two generic parameters. Hence, these variability parameters can be scaled up or down in terms of number of processes and the processing and setup times to fit any process plan.

The main characteristics of process planning are (Bauer et al. 1994, Scallan 2003):

i. it provides the type of processes sequence, i.e., sequential, parallel, etc.;
ii. process planning controls the uniqueness of the final product;
iii. total system loss in operations, i.e., blocked and starved processes or queued work items, and the throughput rates during actual production cannot be determined since process planning is concerned with the operation processes for a single product only;
iv. operations list is based on the available equipment and machines in the manufacturing plant, however, new equipment or machines can be suggested, if they are required or they will improve the quality of the product; and
v. level of details varies based on the nature of the manufacturing or production environment. For example, standard metal-forming operations do not require detailed process planning, however, for highly customised products, high level of details is important for accurate process planning.

Since throughput times of an actual manufacturing system cannot be obtained from the process planning alone, an accurate evaluation of the effect of process variability on the production performance targets is not possible with process planning only.

2.2.2 Production Planning

Production planning is the scheduling of products in terms of quantity and time based on the available resources to meet the customer deadlines. Production planning translates the process plans of different products into a master schedule taking into account the product delivery dates and the availability of resources (Panneerselvam 2012).

Process planning can be used to estimate the required total processing time to complete a single product. On the other hand, production planning can be used to calculate the throughput times for a quantity of different products in an actual production setup. In general, process planning can be considered as a stage of the production planning (Scallan 2003).

2.3 Evolution of Manufacturing Systems

Shipp et al. (2012) highlighted that the trendline of manufacturing systems have seen an exponential increase over the last decades and it is expected to continue for the next 10 to 20 years. These changes were mainly influenced by two main factors; the technological progress and the changes in customer behaviours (Chituc and Restivo 2009). Several other sub-factors were generated from these two main factors, such as the changes in manufacturing strategy, management and methods (Oztemel 2010).

Customer behavior has played a major role over the last few decades to shape the manufacturing systems of today as follows:

i. the production strategy has evolved greatly over the last decades to incline more towards the Make to Order (MTO) rather than the Make to Stock (MTS) strategy (Soman et al. 2004);
ii. different technological approaches were developed to cope with the variability in the finished products without compromising quality and with minimal increase in the capital and operating costs (Pine 1993) (Shipp et al. 2012); and
iii. management strategies have changed from mass production of a single product to more customer-oriented strategies that produce variety of finished products to suit customer needs, e.g., lean production (Paolucci and Sacile 2004).

Jaikumar (1993), Mehrabi et al. (2000), Chituc and Restivo (2009) and Shipp et al. (2012) followed different approaches in identifying the main paradigms for manufacturing systems. In general, three main milestones played a major role in shaping today's manufacturing systems as follows:

i. customisation of products;
ii. conversion to lean and agile; and
iii. introduction of Artificial Intelligence (AI) and Autonomy.

2.3.1 Mass Customisation Era

According to Zhang (2009), mass customisation had replaced the traditional mass production and it is expected, on a large scale, to become the dominant manufacturing strategy for the 21st century. Mass customisation can be defined as a supply chain strategy where manufacturing plays a major role, where products are personalised according to the market needs which is emerged from the customer tendency (Chandra and Kamrani 2004). With the growing competition in industry, manufacturers become more concerned in offering varieties of each of their products in pursuit of satisfying customer needs. However, in order for this strategy to survive, the final customised product cost needs to be as close as possible to that of the mass production (Caster Concepts 2012).

The idea of mass customisation strategy came after the development of new flexible approaches in manufacturing systems that took into consideration the product complexity and the ability to manufacture a variety of finished products. The design and layout of manufacturing system should be able to efficiently handle these challenges while maintaining quality and cost (Beaty 1996). Mass customisation is based on the integration of the global market concept of mass production and the build-to-order concept that was dominant in the pre-industrial revolution era (Davis 1989).

Mass customisation can follow two interpretations for customer needs. It can conservatively be defined as the production according to the requirements of a specific customer; hence, production volume is usually low to medium in this case. More generalised approach is that the customer needs refer to all customised options given to the customer according to a market research of the current customer tendency (Silveira et al. 2001).

In practice, mass customisation usually integrates both interpretations based on the nature of the product. Customer-specific production is applicable to some products, e.g., race cars, and large-volume multi-variety production is a better approach for other products, e.g., computers. The successful implementation of mass customisation is through a balanced integration of these two concepts while maintaining a standardised process and production planning (Silveira et al. 2001).

Several manufacturing systems were developed to achieve this aim. To the extreme of low-volume customer-specific production, cellular or flexible manufacturing systems are the optimal solution, where a small quantity of customer-specific products can be efficiently managed within flexible work cells (Chryssolouris 2006). Each work cell consists of flexible machines applicable for a specific product family (Djassemi 2005).

For the more general case of higher volume production of complex multi-variety products, two approaches were taken. The first is to adapt the machines or processes to be flexible enough to produce multi-variety of products, i.e., groups of general-purpose machines of a certain type, e.g., mill machines, compiled together in a workstation and the work items flow between the workstations (Mukhopadhyay 2015). This type of manufacturing systems is called '*Job Shop*'. On the other hand, this comes at the cost of production planning standardisation since the routing of each product family is stochastic which increases the work in progress (WIP) and limits the production volume (Chryssolouris 2006). The other manufacturing system type, i.e., '*Flow Line*', emphasised the process and production planning standardisation by allowing all products to follow the same routing throughout the production. Hence, flow line are principally arranged in a one-direction flow of materials, i.e., serial flow line, however occasionally, parallel processing may be used and the flow line might have a reverse route for rework (Buzacott et al. 1993, Li et al. 2013). Since the Industrial Revolution and the tendency for mass production, serial flow lines have become widely used and increasingly replaced other types of manufacturing systems (Sennott et al. 2004). However, this manufacturing system is restricted in terms of the range of products it can handle (Mukhopadhyay 2015).

A subtype of flow lines called '*Flexible Flow Line*' was introduced to cope with this challenge as a trade-off between production standardisation and product-variation capability.

This manufacturing system allows flexible machines or processes to produce large-volume medium-variety products without compromising the process and production planning standardisation, i.e., using the same processing sequence. Flexible flow lines have the flexibility to choose between flexible machines or processes in each processing step (Quadt and Kuhn 2007).

In essence, flexible flow lines combine the advantages of both '*Flow Line*' and '*Job Shop*' (Sankar et al. 1997). Hence, flexible flow lines are the main subject of this research. However, variability in processing and setup time of complex multiple products still remains a challenge, especially with the added degree of freedom of flexible machines or processes.

2.3.2 From Mass Customisation to Lean

Lean production represents a prime milestone in the evolution of manufacturing systems (Cappozi and Sacco 2013). Introduction of lean in manufacturing plants had led to a great transformation in facilities layout, organisation structure, manufacturing strategy and process and production planning. These changes have led to a performance peak in terms of cost savings, quality of finished products and on-time delivery of products to customers (Industrial Technology Centre 2004).

Taiichi Ohno, Founder of Toyota Production System (TPS) the cornerstone of lean, has summarised their management strategy as *"All we are doing is looking at the timeline from the moment the customer gives us an order to the point when we collect the cash. And we are reducing that time line by removing the non-value-added wastes"* (1988 cited in Liker 2004). Womack et al. (1991) defined lean accordingly as the elimination of non-value-added activities or wastes and Liker (2004) defined its aim is to "*give the customers what they want, when they want it, at the highest quality and affordable cost*". In general, lean is a customer-oriented manufacturing strategy that strives to ensure customer satisfaction of the final product in terms of value, cost, quality and delivery on time while reducing the capital and operational expenditures of the company by controlling the variability in production processes and efficient utilisation of the existing resources and assets.

Mass customisation introduced changes to the production and process planning of manufacturing systems from mass production with the increased variability. Although variability is important to produce customised products, it can, if inappropriately controlled, lead to an increase in non-value-added activities shown in Figure 2.2.

Mass customisation concentrated on one type of operational waste or non-value-added activity, i.e., overproduction, through customisation of the products according to the customer needs. Lean, on the other hand, extended this concept by relating customer needs to all activities within the manufacturing system, therefore, identified seven more sources of waste. Wastes, as identified in TPS and extended by Liker (2004), are as follow:

i. **Overproduction (Push System):** producing products without a direct order from the customer or a reliable market research (Silveira et al. 2001);
ii. **Waiting:** idle workers waiting for raw materials, tools, equipment repair, etc.;
iii. **Transportation:** internal transportation of raw materials, unfinished products or finished products to and from queues or storage areas. Additionally, external transportation from suppliers to the production plant;
iv. **Overprocessing:** taking longer time or effort than needed to process a product;
v. **Inventory:** Storage of raw materials, queues for unfinished products (WIP) and storage of finished products;
vi. **Motion:** movements of workers to do a task other than product processing, e.g., looking for tools, stacking parts, inspection, etc.;
vii. **Defects:** production of defective finished products subjected to rework or scrap; and
viii. **Underutilisation:** Employees' unrevealed potential skills because of lack of motivation, inspiration or training.

In summary, lean production tweaks the process and production planning so that production is a series of continuous value-added activities that starts from the customer order or potential justified need and ends with the delivery to the customer at the right time, quality and quantity.

Hence in practice, lean process and production planning mitigates the negative effects of increased intra- and inter-variability of the processing and setup times and maintains steady performance targets by limiting the risk of non-value-added activities (Figure 2.3). However, to reach to this objective, a link between the sources of non-value-added activities, i.e., the intra- and inter-process variability within the flexible flow line and the performance targets needs to be established.

2.3.3 Intelligent and Autonomous Lean Enablers

Lean production has emphasised the '*controlled throughput rate*' system-level performance objective through '*reduced system loss*' and '*efficient process and resource utilisation*'. Lean implementation can be done through simple solutions such as labeling and relocation of tools near the relevant process to reduce '*motion*' waste (Basu 2009). However, full lean implementation requires monitoring the activities to determine if non-value-activities, e.g., waiting, starts to arise and perform lean assessment on the available options to mitigate this waste without creating a new waste, e.g., over production.

In other words, lean provides controlled variability effect by decreasing non-value-added activities at the process level as a standard for system-level performance improvement in manufacturing systems. This lean advantage can be enabled, especially in dynamic complex manufacturing setup, using advanced technology (Theuer et al. 2013). Ulrich and Probst (1988) defined complexity as "*a system feature where the degree depends on the number of elements, their interconnectedness and the number of different system states*". The first two can be related to the system itself while the third one is more related to the dynamic nature of the inputs and outputs to and from the system (Scherer 1998). In general, complexity of manufacturing systems is a representation of the variability of parameters, in its general sense, including physical and non-physical elements associated with the manufacturing system.

Intelligent and autonomous solutions can be used to control the value-adding and non-value-adding effects of variability in manufacturing processes autonomously through a control mechanism. The ability of autonomous control systems to take lean decisions on their own allows the production to be in line with the customer requirements and needs (Gronau 2012).

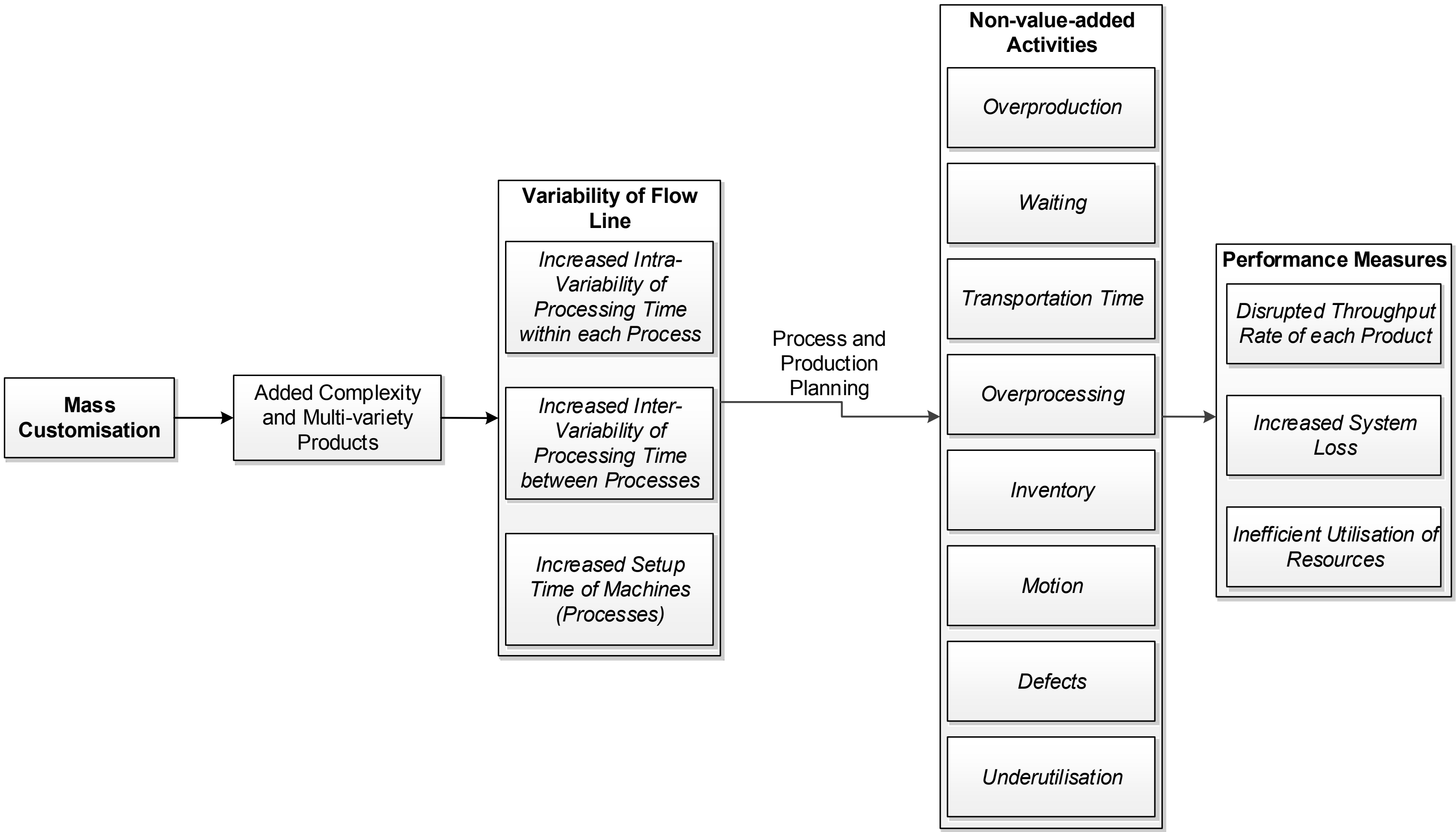


Figure 2.2: Impact of Mass Customisation

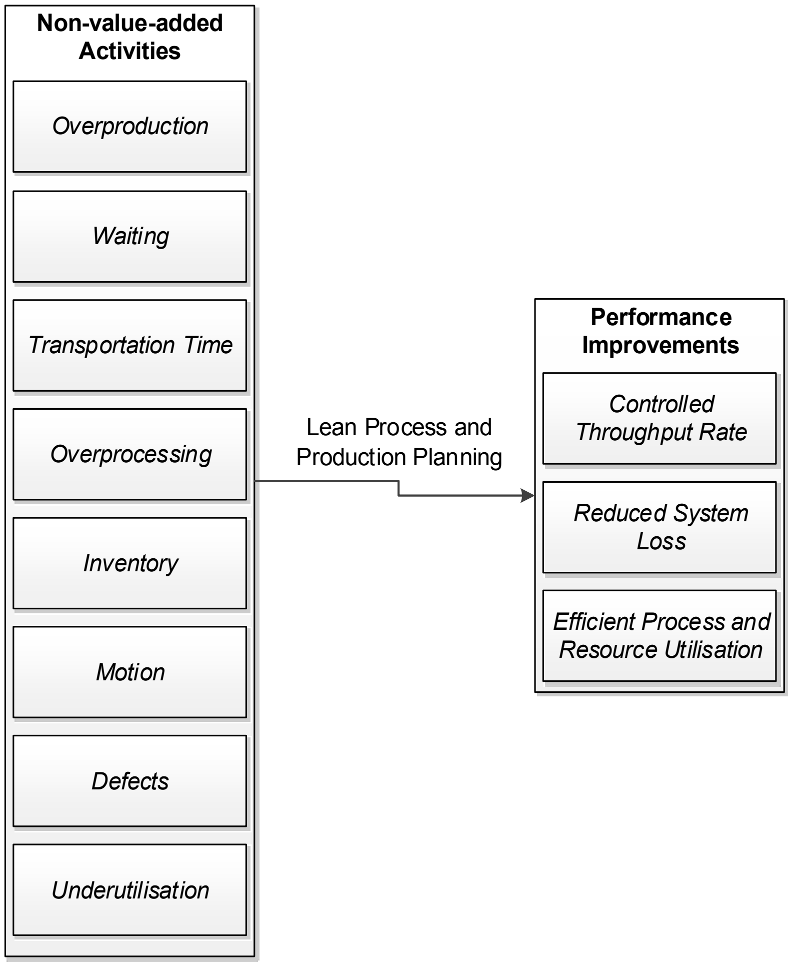


Figure 2.3: Performance Improvements of Lean Process and Production Planning

2.4 Variability in Flexible Flow Lines

Variability can be defined as the change in the effective process and interarrival times at individual process as a result of their stochastic nature which might be related to normal causes, e.g., machine lubrication, age, and resource skills, level of attention, etc., or random events such as machine breakdowns (Hopp 2008, Etman and Rooda 2000). Interarrival time is the time between subsequent work items to arrive at a process. Effective process time is defined as the time the work item spend to and at a process to become ready to be sent to the succeeding process (Jacobs et al. 2003). In other words, the effective process time includes the value-added processing time, when the process is efficiently transforming the work item, and the non-value-added times, where the work item is waiting, being in-transit, overprocessed, etc.

Variability can be introduced in the flexible flow line due to:

i. production of customisable products according to customers' demand, where a single flow line produces different options and features of a product (Leu et al. 1996);
ii. sudden interruptions to the flow line itself such as breakdowns and the change from a product to another, i.e., setup time;
iii. constraints and differences between related products, e.g., 8GB and 16GB flash memory, and their associated processes (Buhne et al. 2005); and
iv. natural reasons, i.e., natural variability, such as friction between mechanical parts, wear, lubrication, etc. (Hopp and Spearman 2011).

Hopp and Spearman (2011) categorised the variability of the flow line based on the coefficient of variation c as low $(c <= 0.75)$, moderate $(0.75 < c < 1.33)$ and high $(c >= 1.33)$.

The classification here is based on the non-value-added activities such as setup and breakdowns. The first category is when no setup is needed and machines or processes are not interrupted by failures. The second for short and the third is for long non-value-added activities at the process such as setups and breakdowns respectively. As mentioned in Section 2.3.1, flexible flow lines are aimed for mass production of moderately variable products, hence, variability will fall under category one and two. However for human-dependent processes category one is more relevant, thus, a good approximation is to consider $c <= 1$. Li et al. (2009) described the following distinguishing key parameters that describe the top level variability of the flow line:

i. **Reliable and Unreliable Machines:** Processing elements with defined efficiency, e.g., 100% for reliable machines, based on the Mean Time to Repair (MTTR) and Mean Time to Failure (MTTF);
ii. **Finite and Infinite Queues:** Storage area in between processes with a restricted or non-restricted capacity of WIP;
iii. **Constant and Variable Intra Processing Times:** The time taken to process the work item can be fixed or not, usually described using a probability distribution profile in case of variable processing times; and
iv. **Homogenous and Inhomogeneous Inter Processing Times:** The variability of processing time from one process to another along the flow line can be zero, i.e., homogenous, or changing, i.e., inhomogeneous.

It is evident that these four key parameters will generate changes to the main two parameters of variability, namely effective process and interarrival times. The research is focused on human-dependent processes. Hence, the machine reliability and setup and queue capacity are not of a concern as presented in Section 1.5.2. In this case, the non-value-added times activities, e.g., waiting, presented in the effective process time are primarily due to the parameter two and three, i.e., intra- and inter-variability of processing times.

The normal distribution tends to be the most applicable form of distribution pattern that represents the variability of human-dependent activities (Martin and Bridgmon 2012). This distribution is also compatible with the case study used for validation, thus, this type of distribution was adapted in this research. However, Li and Meerkov (2009) demonstrated that for non-exponential flow lines with coefficient of variation ≤ 1, such as in the case of this research, the throughput rate is not as sensitive to the distribution type as the coefficient of variation c.

The probability density function (pdf) for normal distribution is given by:

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]} \tag{2.1}$$

The pdf when applied to the flow line shown in Section 1.2, it represents the distribution of intra-variability of each process and is a function of the mean processing time μ and the standard deviation σ. The variability at each process $P_i, i = 1,2,...,N$ is a function of these two parameters in addition to the length of the flow line N.

Since for serial flow line, the arrival rate between a process and another equals the local throughput rate TR_{i-1} (Hopp and Spearman 2011), interarrival time adjustments can be considered as an additional process within the flow line (Wang et al. 2014). The additional process here can represent the demand rate at the entry source or relevant process, or infinite queue, given the assumption of saturated flow line. Processing times for the additional process resemble the interarrival times.

Etman and Rooda (2000) highlighted a limitation with the normal distribution that its range is between minus to plus infinity and processing times cannot be negative. Therefore, they suggested using gamma or negative exponential distributions instead.

However from this research perspective, though normal distribution can go to negative infinity it has less probability to be negative if the coefficient of variation was defined accurately. On the other hand, exponential distribution does not have this privilege since it is defined by one parameter only, i.e., the mean processing time μ which makes it unsuitable to define processes with low or high variability since coefficient of variation is always 1. Furthermore, normal distribution gives the highest probability to the actual mean and distributes the rest of bell shaped probability equally based on c which is a good approximation to human-dependent processes without outages (Hopp and Spearman 2011) than setting the highest probability to zero and distribute the probability exponentially towards plus infinity such as in gamma and negative exponential distributions. A solution for negative processing times can be done by changing the support to $\mu \in (0,\infty)$ enforcing the probability density to be zero when $\mu \leq 0$. Furthermore, both gamma and negative exponential accepts two scenarios that are not realistic $\mu = 0$ and $\mu = \infty$, in fact both distributions give the highest probability density to $\mu = 0$ when $c = 1$ for negative exponential and when $c \geq 1$ in case of gamma distribution.

2.5 Production Performance

The intra- and inter-variability of the processing times within a flow line are transferred to disruptions of the performance targets of an organisation. Carrascosa (1995) stated that the production variability could fluctuate around 30% of the mean system production. Jacobs and Meerkov (1995) has defined a system control property for manufacturing processes called "*property of improvability*". The property is a link between the controllable variability parameters in manufacturing processes and the performance measures targeted for improvement. Fry and Cox (1989) identified three global performance measures; throughput rate, raw material cost and operational cost while the performance measures related to an individual element of the flow line, e.g. process, department, location, resource, etc., are referred to as local. From the lean perspective, the main performance targets are related to three aspects; cost, quality and delivery of the product. Cost and quality are not aimed in this research while the delivery aspect is the main concern.

Windt and Becker (2009) identified four performance targets from the operations perspective in process and production planning as follows:

i. due date compliance;
ii. throughput;
iii. resource utilisation; and
iv. work in progress.

From these two perspectives and restating the performance measures in more generic terms according to the research scope, the research primarily focused on the lean delivery performance target ‘*throughput rate*’ and secondarily on the following lean local performance improvement targets:

i. queue time; and
ii. resource utilisation efficiency.

It is worth noting that these two local performance measures are related to the global performance target, throughput rate as described in the following sections.

2.5.1 Throughput Rate

Steady state throughput rate of the flow line received the most attention in the research concerning evaluative modelling of the performance of flow lines in comparison to other measures (Ambani 2011).

Throughput is defined as the number of completed products, i.e., output. A commonly used performance measure to represent throughput at steady state is the throughput rate defined as (Li et al. 2009):

$$TR = \lim_{t \to \infty} \frac{T_{(out)}(t)}{t} \tag{2.2}$$

where

$T_{(out)}(t)$ is the quantity of products out of the last process in time interval $(0, t)$; and

t is time.

2.5.2 Queue Time

Queues provide a kind of absorption to variability by reducing the blocking of processes (Muth 1987).

At the process level, the relationship between the queue time and throughput rate can be implied from Kingman's equation (Hopp and Spearman 2011, Jacobs et al. 2003):

$$QT_i = \left(\frac{u_i}{1-u_i}\right)\left(\frac{c_{i-1}^{\ 2}+c_i^{\ 2}}{2}\right)\mu_i \tag{2.3}$$

where

QT_i is the queue time at process i;

u_i is the utilisation of process i; and

μ_i, c_i is the mean and coefficient of variation of the processing time at process i.

It is clear that at a process level, an increase to the second and third terms, which represent the intra- and inter-variability of the process, will lead to a reduction of the throughput rate. Hence, the throughput rate and queue time are inversely related.

2.5.3 Utilisation Efficiency

This research focuses only on manufacturing systems with available resources at the respective process or fixed production where enough workforces are hired to fulfill customer orders, hence, no resources starvation (Section 1.5.2).

This performance measure was targeted from the perspective of the efficient utilisation of these resources to perform the processes.

Hopp and Spearman (2011) defined the resource efficiency as:

$$\eta_u = \frac{1}{N}\sum_{i=1}^{N}\frac{TR(i)}{TR^{ideal}(i)} \tag{2.4}$$

where

η_u is the utilisation efficiency;

N is the number of processes;

$TR(i)$ is the local throughput rate of the process i; and

$TR^{ideal}(i)$ is the ideal throughput rate of the process i excluding the variability effects.

It can be observed from Equation 2.4 that the increase in the throughput rate TR will improve the utilisation efficiency.

2.6 Current Evaluative Models of Flow Lines

AI is similar to human intelligence; it develops over time before it can exhibit a form of intelligence. Autonomous machines or processes can have a training mechanism to learn the effect of variability based on local information and past performance and use this to control the system, which will be explained later in Chapter 3. However, building this relationship between local level process variability and system level performance targets in an evaluative model can be advantageous to the optimisation (Spinellis and Papadopoulos 2000) or autonomous control (Zeng et al. 2009).

Therefore, researchers have proposed different approaches to model flow lines to analyse the impact of process variability on the performance measures (Carrascosa 1995, He et al. 2007, Lagershausen and Tan 2015, Li et al. 2009, Papadopoulous et al. 2009, Li et al. 2013, Tan et al. 2015).

Research in flow line modelling dates back to the 1950s. Literature is divided into two main categories; evaluative and generative models (Spinellis and Papadopoulos 2000). Evaluative models are developed primarily for flow line analysis while generative models determine the optimal settings of flow line to satisfy the objective given the constraints on the system (Papadopoulos et al. 2009).

For evaluative models, the modelling approaches can be divided into two main categories; exact state-based Markov analysis for relatively small flow lines with small queue capacity (Papadopoulos et al., 2009) and approximation models for flow lines with an arbitrary number of processes and queue capacities.

Markovian models were widely used by researchers to mathematically model the stochastic variability in flow lines. Markov analysis is a state space model which provides the exact solution as a transition from a state to another with an exponential or phase-based probability over a finite time interval (Norris 1997).

Hunt (1956) was the first to develop an analytical technique for Markov process-derived equations relating processing time with the throughput rate for a flow line with three processes. Miltenburg (1987) used numerical approach to determine the variance in the throughput rate of a flow line with two processes a single finite queue due to interruptions in the flow line in infinite time. Gershwin (1993) used Markov chain to determine an accurate formula to correlate the variance of throughput rate, in a single process flow line, to the process interruptions, e.g., breakdowns.

Queues were represented in the developed analytical formulas by the probability of their effects. Carrascosa (1995) developed analytical formulas for the mean and variance of the throughput rate of a flow line with two processes in relation to the steady state probabilities of sudden interruptions, i.e., machines downtime, within the flow line and changes of the queue capacity. He et al. (2007) extended the approach to include arbitrary number of processes. Wang et al. (2014) and Kang et al. (2015) presented an interesting Markov chain-based analytical model to obtain a closed-form formula for the throughput rate of short exponential flow lines. Lagershausen and Tan (2015) used continuous Markov chain to model the inter-dependencies between processes of a closed-loop flow line with phase-type distributed processes and finite queues to determine an exact solution of interarrival times using numerical iteration.
The Markovian model gives the throughput of flow lines based on the following parameters; number of processes, queue capacity, number of up states at each process, number of down states at each process and the mean processing, repair and failure rates. It is worth mentioning that the processing, repair and failure rates have to follow the exponential distribution for the process to be Markovian and produce linear homogenous equations which can be solved either analytically or numerically.
The main limitations of this approach, especially the first point which precludes the use of this method in this research, can be summarised in:

i. limited to exponentially distributed processing times only;
ii. computational intensive with growing number of states s with the number of processes within the flow line. The number of states with N processes and $N-1$ queues with a capacity C can be determined from (Carrascosa 1995):
$$s = 2^N \prod_{k=1}^{N-1} (C_k + 1) \tag{2.5}$$
e.g., $s = 362{,}544$ states for $N = 6$ and $C = 10$;
iii. suitable for small flow lines only. For long flow lines $N > 6$, equations become very complex and cannot be solved using any analytical or numerical methods (Papadopoulos et al., 2009);
iv. produces equations that are difficult to interpret and understand the causal relationships between variability and throughput rate they represent; and
v. accuracy reduces significantly with increased C (Carrascosa 1995).

Approximation methods based on queuing networks were the mainstream for analysis of larger flow lines. The three main approximation methods are Meerkov aggregation method (Jacobs and Meerkov 1995, Li and Meerkov 2003, Li and Meerkov 2009) and Gershwin decomposition method (Gershwin 1994). The first follows backward and then forward aggregation to approximately convert the long flow line into a two processes one queue flow line, where Markov analysis can be applied. The same concept applies to the decomposition method but instead of aggregation, the flow line is split into a combination of two processes one queue flow lines with an equation for each. Dallery-David-Die algorithms were developed by (Dallery et al. 1989) to solve the decomposition equations. Li et al. (2009), Papadopoulous et al. (2009), Li et al. (2013) and Tan et al. (2015) provide a comprehensive overview and illustrate the latest developments in the approximation methods for the throughput analysis of flow lines. These methods converge and usually produce accurate results, however, they assume deterministic, exponential or phase-based distributions of processing, repair and breakdowns (Enginarlar et al. 2006, Li et al. 2013).

Simulation has been widely used to estimate the performance of a flow line that is complex or impossible to be modelled mathematically (Brandimarte and Villa 1999). It also allowed analysis of different distribution patterns other than exponential which allowed more flexibility in the representation of actual flow lines (Enginarlar et al. 2006). On the other hand, simulation evaluative models are case-specific and time-consuming to build the simulation model.

Another approach is to carry out an empirical and analytical study to produce closed-form formulas to represent the performance of flow lines (Papadopoulous 1996). Simulation and data mining have been the main drivers for the empirical approach (Papadopoulos et al. 2009) while analytical formulas were derived using first principles modelling such as queuing theory (Hopp and Spearman 2011) and holding time model developed by Muth (1987). The developed formulas describe the effect of variability on several performance measures such as throughput rate, work-in-progress, blocking, starvation, system delay, etc. Unlike Markov analysis, these formulas are not exact, however, they can give insights into the system behavior and help with process improvement. This approach gives more flexibility in the distribution used for the process variability, timesaving than simulation evaluative models and fast offline analysis without disruption to the actual flow line in study.

In general, the main advantages of this technique are:

i. applicable to exponential and non-exponential distributions of process variability;
ii. simple and relationships can be easily understood;
iii. can be used to optimise the planning and operations of flow lines;
iv. if accurately tested, can provide a reliable model close to exact mathematical models, e.g., Blumenfeld (1990) generated an error of ±1-5% when compared to the exact solution using Markov analysis (Hillier and Boling 1967);
v. computation easy and can be implemented with any programming language; and
vi. time taken to process these simple formulas is less than processing complex numerical solutions or building simulation models.

On the contrary, the main limitation of the empirical formula is that they are not mathematically proven. However, since the current first principles models to provide a standalone solution, require the process variability to follow deterministic, exponential or phase-based distributions (Slack and Wild 1980, Enginarlar et al. 2006, Li and Meerkov 2009, Meerkov and Yan 2014), the empirical approach offers, if carefully tested and validated, a good alternative route.

2.6.1 Empirical Formulas for Synchronous Flow Line

Barten (1962), Anderson and Moodie (1969), Knott (1970) Buxey et al. (1973), Slack and Wild (1980), Medhi (1991), Khalil (2005) and Hopp and Spearman (2011) used data mining combined with theoretical analysis or simulation data to investigate a formula for system loss-based performance measures such as mean system delay, work in progress, optimal queue capacity, queue time and blocking and starvation of each process.

As for the evaluative modelling of the throughput rate of synchronous flow lines, Muth (1987) built a formula of the throughput rate using data mining and theoretical analysis. The formula was tested on a flow line with no queues and two to ten processes with exponential, Erlang, uniform and fixed distribution types of the intra-variability of the individual process. Blumenfeld (1990) extended Muth (1987) formula using analytical analysis to include normal and binomial distributed processes, longer flow lines and with queue capacity up to 10 work items.

Blumenfeld and Li (2005) developed an analytical closed-form formula for the throughput of synchronous flow lines with deterministic processes and exponentially distributed failure and repair rates. Blumenfeld (1990) formula is given by:

$$TR = \frac{1}{\mu\left[1 + \frac{1.67(N-1)c_{av}}{1 + N + 0.31c_{av} + 1.67NQ/(2c_{av})}\right]}, \text{ where } c_{av} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} c_i^{\,2}} \qquad (2.6)$$

2.6.2 Empirical Formulas for Asynchronous Flow Line

Papadopoulos (1996) used the holding time model (Muth 1973) to develop a formula for the throughput rate of synchronous flow lines with N exponentially distributed processes and no queues. The formula includes coefficients that need to be first solved numerically in order to obtain the throughput rate.

It is evident from previous research (Li and Meerkov 2005, Li and Meerkov 2009, Meerkov and Yan 2014) that for asynchronous non-exponential flow lines, the process with maximum mean processing time $\mu_{\max}$ plays a major role in determining the throughput rate of flow lines. Hence, modelling of TR using the distribution parameters μ_i and $\sigma_i, i = 1,2,...,N$ across the flow line length N is not an appropriate method. In fact, the throughput rate for a serial flow line with deterministic processing time of each process (i.e., TR^d such that $c_i = 0, i = 1,2,...,N$) can be obtained as a function of $\mu_{\max}$ only using the following formula (Li and Meerkov 2009):

$$TR^d = \frac{1}{\mu_{\max}} \qquad (2.7)$$

The formula implies that the process with the maximum processing time controls the throughput rate of the flow line. However, introduction of intra-variability at each process, presented by σ or c, produces more complex changes in the throughput rate that cannot be solely presented by $\mu_{\max}$.

Li and Meerkov (2009) expressed this relationship in the following form:

$$TR = TR^d - \left[TR^d - TR^e\right]c_{av}, \text{ where } c_{av} = \frac{1}{2N}\sum_{i=1}^{N} c_i \qquad (2.8)$$

TR^e is the throughput rate for exponential processes, i.e., $c_i = 1, i = 1,2,...,N$, and c_{av} is the average coefficient of variation for all processes. Equation 2.8 shows that the intra-process variability reduces TR than the deterministic case. The change, i.e., the second term, is proportional to the difference between TR at $c = 0$ and $c = 1$.
Although Equation 2.8 was developed with deterministic processing times and a non-exponential distribution was used to represent the machine reliability only, the formula is still applicable to the opposite case, i.e., deterministic failure rates $\lambda_i = 0, i = 1,2,...,N$ and non-exponential distribution of processing times. However, the main source of variability, i.e., $c_i = 1$, still requires the use of other methods, such as simulation, since as explained earlier in Section 2.6, the Markovian-based state-space approaches to calculate TR^e are not applicable to this research.

2.7 Summary

In general, manufacturing systems become more flexible to produce a range of complex products to suit the customer needs. However, this flexibility, which is necessary to produce customer oriented products, led to an increased variability within the manufacturing system in terms of layout, job routing and processing and setup times of different products. Flexible flow lines present a suitable layout to limit these challenges with the standardised job sequencing. Variability is transformed to fluctuations in local performance measures, e.g., system delay, and subsequently global performance target, i.e., throughput rate. The lean and AI advantage in this aspect is the use of intelligent solutions to control the effect of increased variability through reduction of the non-value-added activities in flexible flow lines. Evaluative modelling is a key part of the control mechanism. Analytical models are not applicable for non-exponential flow lines, hence, the empirical route is a good alternative. Closed-form empirical formulas give extra advantages over simulation such as they are time-efficient, easy to interpret and simple to apply. Next chapter dives into the existing autonomous and optimisation solutions that were developed to control the variability effects on the performance targets specific to research.

3 AUTONOMY IN FLEXIBLE FLOW LINE

3.1 Introduction

Fast paced changes in customer behaviour over the last decades and dynamic pace of today's market has generated the need for flexibility in manufacturing systems to cope with the frequent changes in customer specifications and demand. As shown in Figure 3.1, to cope with these changes, flexibility was built in manufacturing systems, which increased process variability and complexity and challenges in process and production planning. This opened the doors for development of new techniques and methods, e.g., autonomous control, evolutionary algorithms, etc., to deal with the arising challenging in order to achieve performance targets.

However, to maintain the performance targets of the systems, the impact of variability on the performance of the manufacturing system needs to be analysed. Autonomous systems have used different AI techniques and approaches for the implementation of the '*controlled variability effects*' rule that was also stressed by lean production.

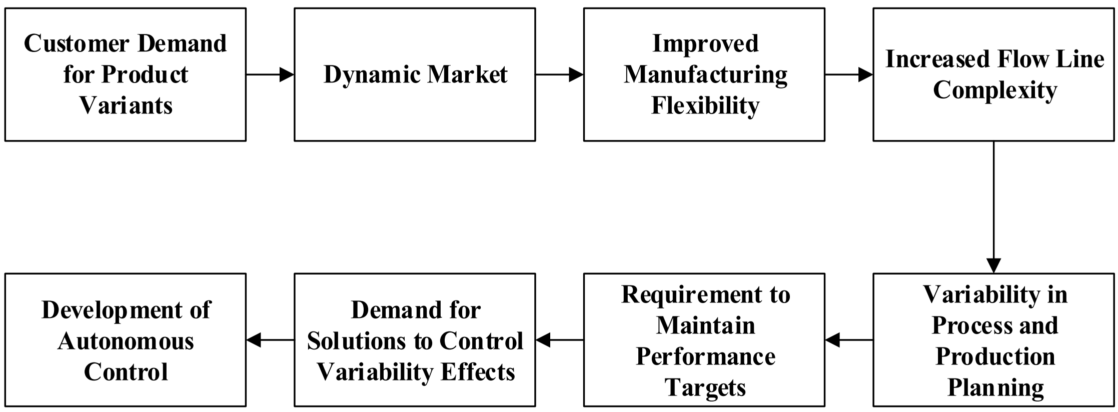


Figure 3.1: Development Cycle of Autonomous Variability Control

In this chapter, a brief outline of the translation of intra- and inter-process variability into flexibility within the autonomous control is explained. Afterwards, an overview of the existing autonomous control methods and techniques for flexible flow lines to manage this flexibility is covered.

3.2 Manufacturing Flexibility

Flexibility in manufacturing systems is a measure of the capability of processes to adapt and the control system to take a different decision in response to changes within the manufacturing system (Baykasoglu and Gocken 2011). To achieve these goals, flexibility in flow lines and planning is one of the key solutions (Heilala 1999). Flexibility is translated in autonomous manufacturing to the ability of the actual processes and assembly lines to adapt to the changes in customer order or market need (Heilala 1999). Additionally, flexibility provides the capability of the hardware, i.e., processes and assembly lines, to be integrated with the adaptive autonomous control software (Windt and Jeken 2010). This area has influenced the provision of some concepts such as Flexible Manufacturing and Reconfigurable Manufacturing Systems (Scholz-Reiter and Freitag 2007). Existing autonomous control systems have been developed for every stage of manufacturing system, listed in Section 2.2, to help in communication between these stages and to automate the tasks in each stage (Oztemel 2010).

Flexibility in manufacturing systems has evolved greatly over the past decades. However, flexibility of current autonomous systems still does not reach to the flexibility achieved by humans and a study suggests that the human involvement in some manufacturing operations, e.g., assembly, is necessary to reach to the optimal level of flexibility and adapt to changes in customer specifications (Bley at al. 2004). The design and choice of autonomous system is a tradeoff between incurred cost and required flexibility, since highly autonomous control will not be cost-effective for simple flow lines with low degrees of flexibility. At the same time, increasing the level of autonomy over a certain level, even for complex systems, can lead to a chaos which will eventually lead to a dip in the performance indicators (Windt et al. 2008). The degree of autonomy incorporated in the system has to take into consideration the degree of complexity and flexibility of the flow line to decide on the optimum cost-effective solution.

Sethi and Sethi (1990) identified three levels of flexibility; component, system and aggregated. Wiendahl et al. (2007) has identified three perspectives to classify manufacturing flexibility; order, product and resource. Windt and Jeken (2009) combined the two concepts and added another sub-category, i.e., allocation flexibility, as shown in Figure 3.2.

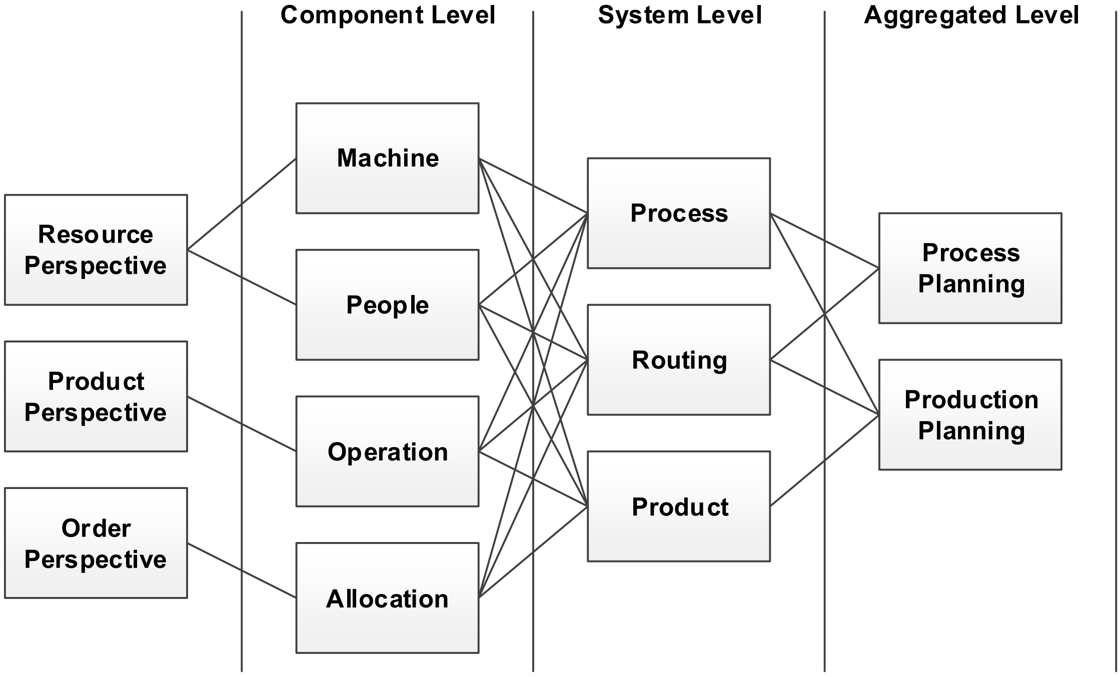


Figure 3.2: Classification of Flexibility (based on Windt and Jeken, 2009)

In general, it can be concluded that the flexibility within a manufacturing system will have multiple processing options or decisions, in terms of product variants and job allocations, for each process which will be accumulated to generate the process and product planning. However, at any point of time the flexibility will generate different possible schedules with various degrees of intra- and inter-variability of the processing times which will eventually transfer to the performance targets. The main task of autonomous control is to use the decision flexibility to control the system variability for the purpose of achieving the required performance targets.

3.3 Manufacturing Intelligence and Autonomy

3.3.1 Conceptual Background

Introduction of Artificial Intelligence (AI) and autonomy in manufacturing systems was one of the major milestones that transformed manufacturing systems. Dagli (1994) suggested that the first integration of AI in manufacturing systems goes back to the late 1980s.

AI is a section of computer science based on the replication of human reflexes and reactions to an input (Nath 2009). Humans behave in an intelligent way to different situations they are facing on daily basis. This human intelligence is mainly based on the accumulation of expertise and knowledge gained during the course of life. Meystel and Albus (2001) defined intelligence "*is to perceive the environment in which the system is operating, to relate events taking place around the system, to make decision about the events, to perform problem solving and generate the respective actions and control them*". Human brain does not very often perform mathematical calculation to reach to a decision. Instead, human depends on their stored knowledge of similar situation or set of situations to decide what to do in a new situation, if full knowledge of it does not exist.

Manufacturing intelligence is defined as the integration of primarily AI with other non-AI techniques, e.g., clustering, into the manufacturing stages, discussed in Section 2.2, for the purpose of increasing their intelligence level (Zhou et al. 2010).

Scholz-Reiter and Freitag (2007) defined autonomy as the "*independence of a system in making decisions by itself without external instructions and performing actions by itself without external forces*". Windt et al. (2008) defined autonomous control as the ability of "*single entities (e.g. parts, pallets, orders or work-station) to render information and to make decisions on their own... by decentralised decision-making in heterachical systems"*. Autonomous manufacturing system can be defined as the manufacturing system, where processes self-optimise, by means of manufacturing intelligence, their decision flexibility to adapt with the dynamic and variable nature of modern manufacturing environment.

AI techniques were widely used in the process and production planning of intelligent and autonomous systems to (Oztemel 2010):

i. adapt to upstream changes in preceding design stage; and
ii. allow flexible flow lines to adapt to downstream changes in customer demand.

Complexity of flexible flow lines is proportionally related to the increase of product complexity. Failure to match the production cycle to the market needs and in sync with the customer demand rates can cause the performance of the organisation to drop significantly and accordingly reduce the customers' satisfaction (Hitt et al. 1998).

Optimisation of flow lines according to the changes from the customers end requires the autonomous system to take the right decision at the right time. To achieve this, the autonomous system has to collect the required information and match it with the required output and take decision accordingly in a very short time (Raol and Gopal 2013). Figure 3.3 represents the autonomous control loop based on that concept.

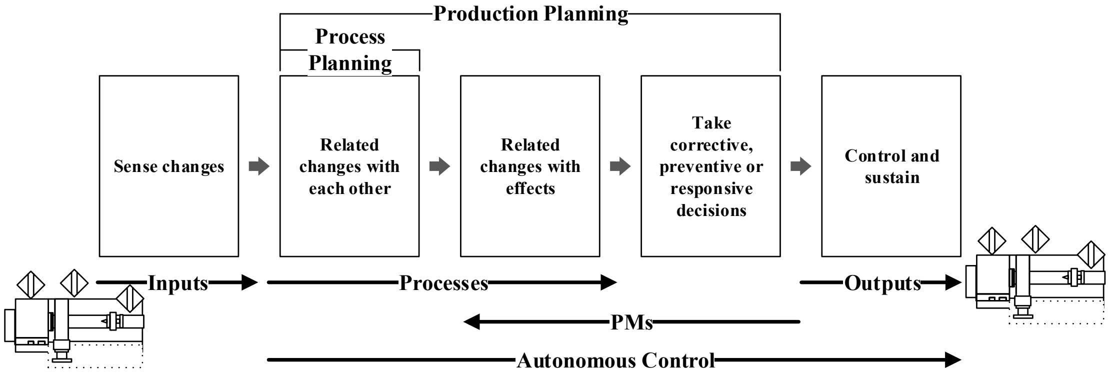


Figure 3.3: Autonomous Control Loop for Process and Production Planning

3.3.2 Intelligent Agents

Intelligent agents is a branch of manufacturing intelligence which is common in large flow lines with high flexibility that requires quick adaptability of hardware and software and direct control of the physical components (Dashkovskiy 2011). In such a system, the manufacturing entities, i.e., machines or processes, within the autonomous control system are referred to as intelligent agents. Intelligent agents control the autonomous flexible components, of each machine, to perform the local goals based on the local information (Oztemel 2010). In essence, intelligent agents represent a combination of manufacturing intelligence and autonomy.

The agent based system was introduced to enable manufacturing systems and assembly lines to handle frequent changes in customer orders more efficiently and to overcome the complexity in centralised control system in earlier manufacturing systems (Scholz-Reiter and Freitag 2007). Intelligent agents based systems are more concerned with decentralising the control of manufacturing systems by dividing the flow line into subsystems containing autonomous entities (Scholz-Reiter and Freitag 2007).

The relationships between manufacturing entities are also constructed between intelligent agents to enable communication and flow of information between them in their local environment (Windt and Jeken 2010). Intelligent agents represent the autonomy, from the software side, for holonic reconfigurable manufacturing systems (Scholz-Reiter and Freitag 2007). Intelligent agents can have a direct interaction with the physical system for update of information.

Artificial Neural Network (NN) and Evolutionary Computation, e.g., Genetic Algorithm (GA), are commonly used computational intelligence methods to provide AI within each holon, i.e., intelligent agent, of the autonomous control system (Oztemel 2010) (Zhou et al. 2010).

The NN predicts future, or possible events, based on historical data. The NN learns by associating measured inputs to immeasurable outputs and then predicts, after a learning phase, the values of immeasurable outputs at any condition of inputs (Fu et al. 2006). This type of AI is based on weighted estimation of the nonlinear relationship between dependent, i.e., output, and independent, i.e., input, variables (Benitez et al. 1997). Rippel et al. (2010) integrated NN into an existing autonomous control decision to help with the job routing decision.

However, the NN has the following limitations:

i. requires training to exhibit intelligence; knowledge is acquired from interaction with the system so it becomes more intelligent with time;
ii. on the long run, NN is subjected to ‘*catastrophic forgetting*’ since the trained decisions based on collected data are getting contradicted by the new data (Date and Kurata 2008);
iii. intelligent as the information fed to it, therefore, it has to be trained properly with sufficient scenarios of the situation to give accurate predictions (Rajkumar and Bardina 2003);
iv. black-box method of control (Benitez et al. 1997) since the training logic behind the NN is not presented to the user; and
v. the nonlinear relationship between inputs and outputs is formulated by hidden nodes (Benitez et al. 1997) and therefore reproduction of results requires implementation of the NN itself.

GA is one of the most popular evolutionary algorithms used in intelligent and autonomous manufacturing systems (Zhou et al. 2010). GA is an optimisation method which follows the following procedure to obtain the best solution for a problem (Oztemel 2010):

i. starts with a set of random possible solutions, i.e., chromosomes;
ii. examines solutions against the problem using a fitness function;
iii. population evolves using genetic operators (mainly crossovers and mutation) to produce a better solution to the problem, i.e., better fitness; and
iv. when no better solution, i.e., fitness, can be produced, it is considered the optimal solution.

GA shares some similarities with NN since both of them become more intelligent over time though with different approaches. GA mostly starts with a population of random solutions and starts to optimise it to get the best fit to the exact solution. NN, instead, does not define random solutions and instead considers these random solutions as input independent variables to the output dependent variable, i.e., objective, and determines the correlations, i.e., weights, between these input and output variables.

Therefore, although GA can start from nothing and reach to an optimum solution, learning period for GA can be long since more complex, i.e., mutated, solutions are generated and examined at each reproduction process.

3.4 System Stability

As mentioned in the introduction, maintenance of the performance measures regardless of the flow line variability is a main concern for any organisation. Stability is an importance performance measure of the system ability to handle dynamic fluctuations in the inputs to the system which will lead to an increase of the mean interarrival rate beyond the stability limit, e.g., seasonal product demand. To reach to the stability state, the arrival rate at a process i has to be less than or equal to the throughput rate of this process (Scholz-Reiter et al. 2005):

$$TR(i-1) \leq TR(i) \tag{3.1}$$

Stability remains a challenge to autonomous systems because decisions are taken at the process level. Evaluation of other performance measures should always be restricted to the stability boundaries (Scholz-Reiter et al. 2005).

3.5 Existing Autonomous Control Techniques

Scholz-Reiter et al. (2010) defined autonomous control as the "*shift of decision-making capabilities from the system layer to its elements*". The elements refer to physical entities in the manufacturing system, e.g., process, product, etc., and non-physical attributes such as local information at each entity, e.g., processing time (Dashkovskiy et al. 2011). One of the main aspects of autonomous control is the degree of freedom of the elements to take process and production planning decisions on their own based on the current state at each moment of decision making instead of depending on a pre-determined schedule (Scholz-Reiter et al. 2009b).

Autonomous elements can take decision based on local information only or they can seek necessary information from other elements in the system (Scholz-Reiter et al. 2010). Regardless, autonomous control should be incorporated into the system to assist in process and production planning without causing any changes to the way the product is processed (Windt and Becker 2009). Therefore, the cycle time of processes will remain the same but the order and process sequences will differ autonomously.

The autonomous control methods can be categorised into rational and bounded rational methods. The first is based on rules that an expert can take with a specific performance-based target in mind, e.g., reduce waiting time; improve due date delivery; etc. (Scholz-Reiter et al. 2005, Scholz-Reiter et al. 2006, Zozom et al. 2003). Another rational method designed for complex and dynamic production environment is called Distributed Logistics Routing Protocol (DLRP) (Rekersbrink 2012). In this method, the orders investigate possible routing alternatives at the beginning of the flow line and then update the routing continuously after each processing step based on the available local information such as processing and setup times, deadlines, etc.

The bounded rational methods are a set of algorithms that replicate the behavioural intelligence of some biological orders, e.g., ants, bees, etc. These methods follow the concept of depending on the past instead of future events to learn and adapt the best routing to be followed depending on stored pervious performance measures of production elements, e.g., waiting, travel, processing times, etc. Different techniques following different biological creatures were developed such as Pheromone Based, Honey Bee Algorithm and Chemotaxis policies (Cirirello and Smith 2001, Tsutsui and Liu 2007, Armbruster et al. 2006, Scholz-Reiter et al. 2008a, Scholz-Reiter et al. 2008b, Scholz-Reiter et al. 2010).

In terms of performance of the autonomous control method, Windt et al. (2010) carried out an interesting study to classify the autonomous control methods based on their performance. The results of this study show two distinctive behavioural classes; one is high and the other is low in performance. Queue Length Estimator falls under the first category and it had shown superior performance over the other autonomous control methods. As for the second category, Past Events Based method showed a poor performance. Both methods fall under the rational category.

The main distinctive characteristics of autonomous control method are (Windt et al. 2010):

i. **Information Source:** the element that shares information rendered in the decision making;
ii. **Information Type:** past, future (predicted) or both;
iii. **Decision Variables:** factors that need to be controlled to reach performance targets;
iv. **Decision Steps:** number of decisions to be taken; and
v. **Algorithm:** the control logic.

3.5.1 Heuristics and Autonomous Control

Use of autonomous control has been associated to the sequencing and assignment of products to processes which is a core challenge in scheduling of flexible flow lines. Grundstein et al. (2015) investigated another flexibility decision that is also important to scheduling, i.e., order release methods, to be integrated with autonomous control and studied the effect of such on the performance targets. Predictive scheduling has been always associated with static deterministic flow lines, where a production plan can easily be decided prior to commencement of the work (van Brackel 2009) so autonomous control application here is limited. With the introduction of variability to this system, autonomous control methods deal with the dynamics of such a system by handing over the decision to the elements to decide based on local information, i.e., reactive scheduling (Kang et al. 2014).

For routing optimisation problem, Scholz-Reiter et al. (2010) investigated the performance of decentralised autonomous control against the centralised heuristic methods. The first control method takes decision based on the current system state, hence, handles dynamic situations more appropriately. The second performs better in a static environment. One of the widely used optimisation package in research and industry is OptQuest (Laguna and Marklun 2013). OptQuest is a meta-heuristic and mathematical optimisation tool that runs under various simulation environments. The metaheuristic search methodology used in OptQuest to optimise the decision variables is based on scatter and tabu methods. Other supplementary methods, e.g., NN and linear and mixed integer programming, are also used to assist the search process (Shortle at al. 2014).

The difference between the heuristics and autonomous control methodologies is how the control system deals with the variability imposed by the dynamic nature of the input, i.e., arrival rate. Autonomous control handles it by measuring the local effects of variability on individual element, e.g., queue length for queue length estimator, while scheduling heuristics targets the system level performance level, e.g., makespan.

Both techniques have different approaches to reach to the best solution with advantages and limitations for each, however, there is always an uncertainty that the best solution is the optimal one. One way to deal with such a problem is to know the effect of each decision on the performance targets prior to taking the control decision. This requires an accurate evaluative model relating variability of the flow line to the performance targets and association of the autonomous control decision to the model.

3.5.2 Queue Length Estimator (QLE)

This method evaluates the waiting and processing times, i.e., workload, of each possible route for the product and follows the shortest route, i.e., shortest throughput time per product per stage. The evaluation process will depend on the queuing parts and the changes in processing time of the same product from one machine or process to another. The evaluation is repeated each time a part leaves a machine or process and before it goes to the succeeding machine or process (Scholz-Reiter et al. 2005). Figure 3.4 shows a demonstration of the evaluation process.

As described, this concept is not different than the DLRP except that DLRP is computational intensive since more information, e.g., customer deadlines, are collected at the process level and used as part of the autonomous decision.

Complexity of DLRP, due to the large amount of information used to reach a decision, is one of the main limitations of the technique. The less complex autonomous controls '*QLE*' provided better performance during implementation for flexible flow lines (Windt et al. 2010).

3.5.3 Past Events Based (PEB)

This method is also used to determine the optimal routing for multiple products within flexible flow lines. The method is based on the previous events rather than future predictions. Recorded performance measures, i.e., waiting and processing time, during simulation are used to make a decision on the job routing. At the exit of a completed processing stage, the product goes to process with minimal historical average throughput time for the respective product type (Scholz-Reiter et al. 2006).

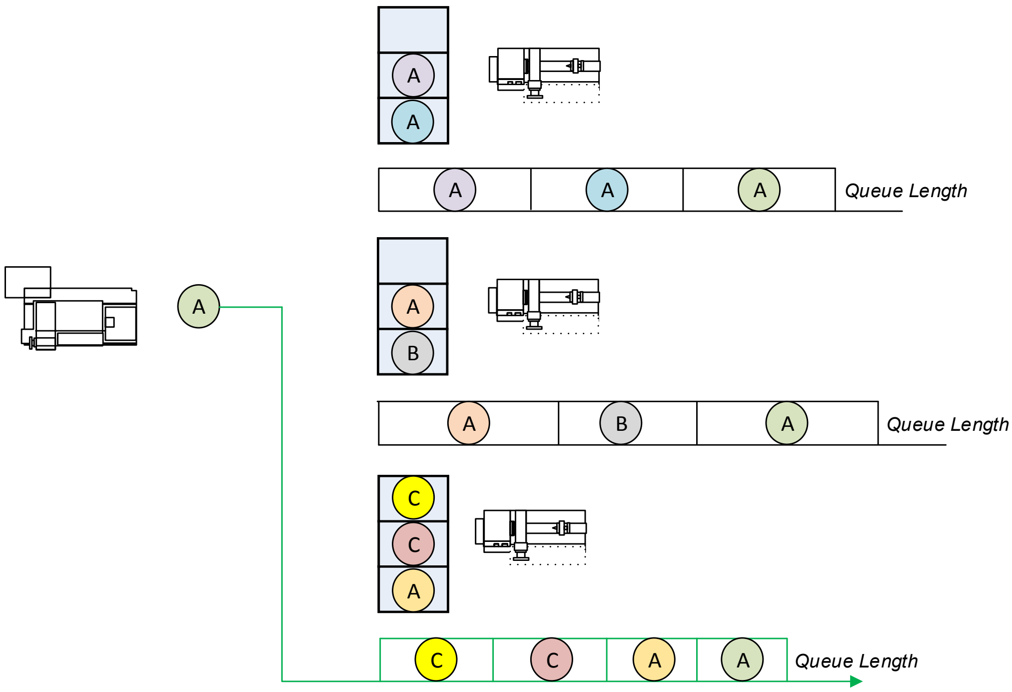


Figure 3.4: Queue Length Estimator Method

3.6 Summary

Practical implementation of the '*controlled throughput rate*' rule, introduced in Chapter 2, is achieved through increasing the degree of intelligence and autonomy within the manufacturing system and improving the autonomous decision flexibility to control the variability within the flexible flow line. Intelligent agents represent an example of the software of such implementation, where the autonomous control decision is in the hands of each individual process instead of a centralised system.

This decentralisation of autonomous decision usually leads to better process and production planning decisions, e.g., scheduling, in a dynamic situation than heuristic optimisation methods, however, the latter is better in static situations. From all autonomous control methods, QLE is one of the best while PEB is among the worst (Scholz-Reiter et al. 2006, Windt et al. 2010).

The following chapter explains the methodology implemented in this research to build the evaluative model and produce its own implementation method of the '*controlled throughput rate*' rule.

4 METHODOLOGY

4.1 Introduction

This chapter gives a detailed description of the methodology used to build an evaluative model of the throughput rate for synchronous and asynchronous human-dependent serial flow lines and apply it for development and validation of an autonomous control method for flexible flow lines.

The research aim is to build a simple and quick autonomous-decision-support mechanism that accurately predetermine the effect of stochastic variability on the system-level performance target of synchronous and asynchronous human-dependent serial flow lines. This empirical research investigated realisation of this aim through two methodological frameworks:

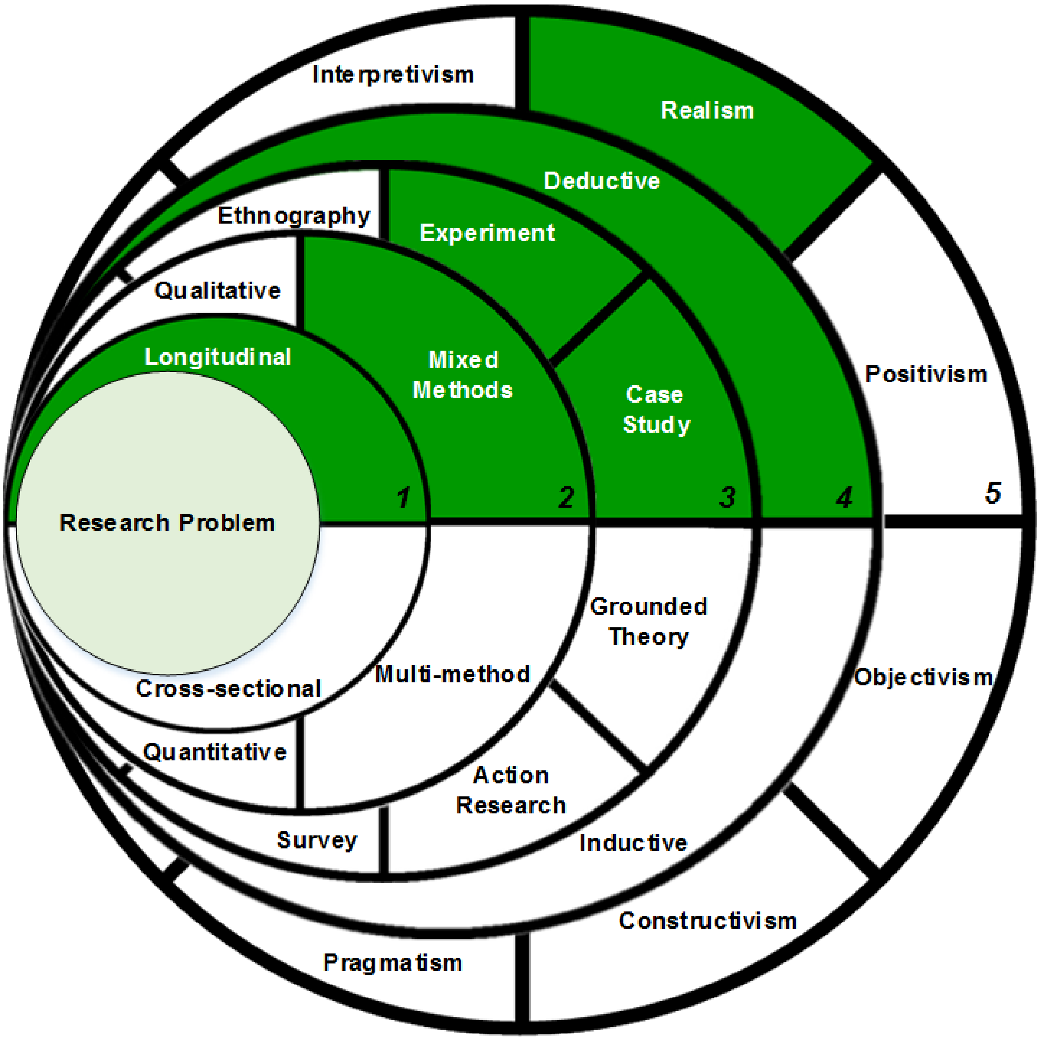
i. Data Mining Framework: a standardised framework that investigates the degrees of freedom in each stage of the mining process to:
   a. generate representative data sets for the intra- and inter-stochastic process variability and use simulation to determine the steady state throughput rate for the variability scenarios with high certainty;
   b. apply statistical analysis to build a generic representation of non-exponential serial flow lines based on the impact of variability on the throughput rate; and
   c. use supervised machine learning methods to examine and select a regression data mining model for the throughput rate of:
      - synchronous human-dependent serial flow lines; and
      - asynchronous human-dependent serial flow lines.

ii. Autonomous Control Framework: uses the empirical formula for autonomous control and examines the performance of this method against other autonomous control and optimisation methods for flexible flow lines.

The chapter starts with an overview of the methodology and frameworks used to approach the research questions stated in Section 1.4. The research problem is then identified in terms of the variables used to build the evaluative model. Afterwards, the chapter gives an overview of the tools used in the investigation. It then moves to the detailed data collection, pre-processing and analysis steps that were undertaken to implement the frameworks. Finally the methods and steps applied to validate the evaluative model and the autonomous control method are stated.

4.2 Research Methodology

The research methodology implemented for this study is presented in the research onion, developed by Saunders et al (2009), as shown in the highlighted green fractions on Figure 4.1. The following two sections will elaborate on the selections for layer 2 and 3, namely research choice and strategy, as selections for other layers are self-explanatory given the aim of this research. Section 4.2.3 is devoted to the methodological frameworks that govern this study.



***1** is the Time Horizon layer;*

***2** is the Choice layer;*
***3** is the Strategy layer;*
***4** is the Approach layer; and*
***5** is the Philosophy layer*

Figure 4.1: Research Methodology using Saunders et al (2009) Research Onion

4.2.1 Research Strategy

Research strategy is the "*the general plan of how the researcher will go about answering the research questions*" (Sauders et al. 2009).

In this research, the strategy is to carry out '*experiments*' to determine the relationships between the independent variables and the dependent variable based on the research design. A '*case study*' is then used to examine and validate the identified relationships. Figure 4.2 gives an illustration of the research strategy.

Two methodological frameworks in pursuit were established, as detailed in section 4.2.3, in pursue of answering the research questions in a systematic manner to create the research outcomes.

4.2.2 Research Choice

Sauders et al. (2009) classified the research choices into qualitative, quantitative, mixed-methods and multi-methods. Clearly, pure qualitative approach does not apply to this research. Multi-method and mixed-methods share the advantage of combining different methods in a single research step, however with multi-method, these methods have to be either quantitative or qualitative (Sauders et al. 2009). Hence, quantitative, mixed-methods and multi-method quantitative choices can apply to this research.

4.2.2.1 Quantitative Research

Quantitative research is concerned with analysing measurable results of variables related to the area of research and reproducibility of the obtained results (King 1994). This strategy fits well with the research problem as measurability and reproducibility are two important aspects to build and analyse the performance of the evaluative model and the autonomous control method.

The number of case studies associated with this research strategy is usually large in order to validate the results and the obtained relationships between variables (Tewksbury 2009).

Advantages of quantitative research include (Tewksbury 2009):

i. production of reliable results based on quantified variables and experiments;
ii. results can be verified and validated by other researchers;
iii. replicable results; and
iv. future works can be identified and carried out by other researchers.

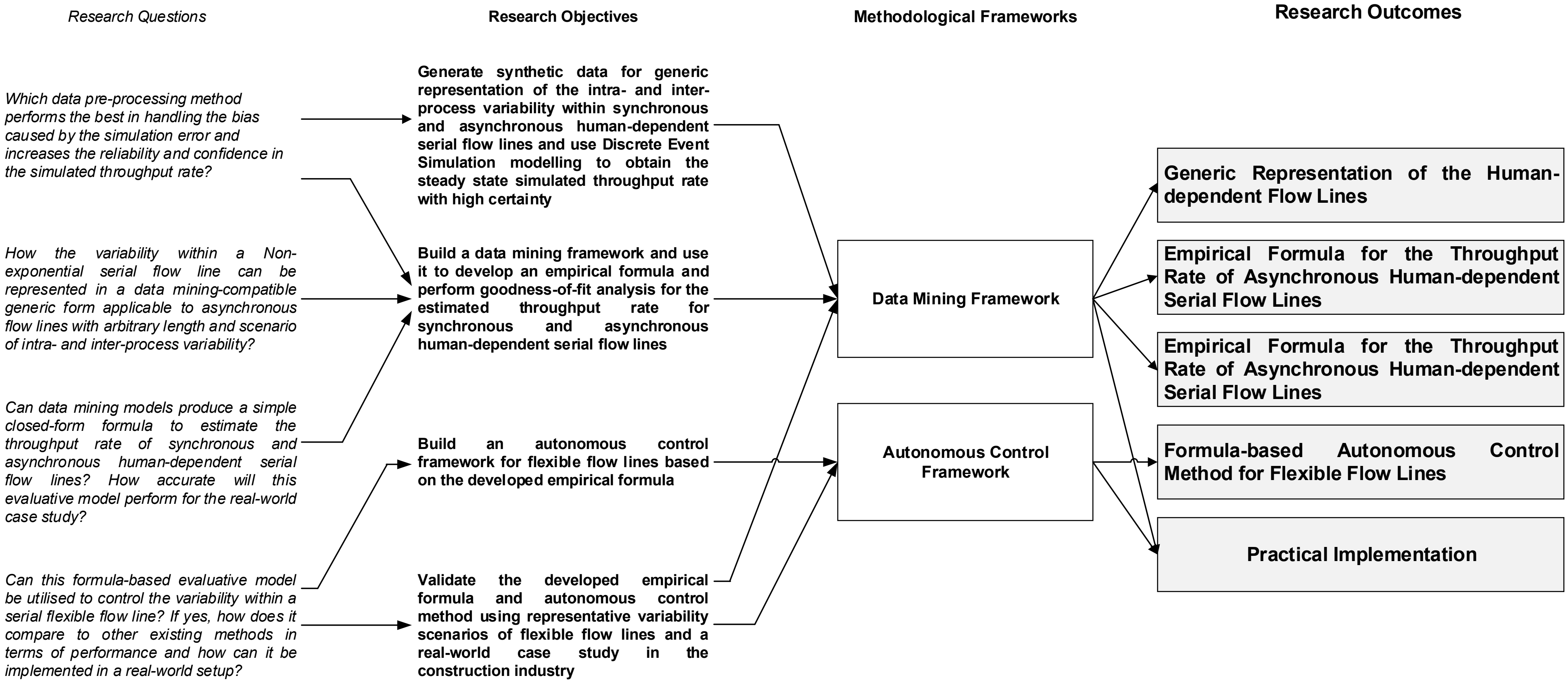


Figure 4.2: Research Strategy

The main drawback of quantitative research is that some variables are not fully measurable or subject to human judgment. Therefore, their relationship with other variables is changeable which questions the reliability of the research outcomes.

Another disadvantage is that the quantitative research cannot be completely detached from the author qualitative values since the research is the outcome of the author's observations based on the search results and the related literature which can include some reasonable assumptions at the time of research (King 1994).

As a result, mixed-methods study is a more appropriate choice for this research than multi-method quantitative as it gives the flexibility of qualitatively and quantitatively evaluate a single research step.

4.2.2.2 Mixed-methods Study

Mixed-methods or triangulation of quantitative and qualitative methods can be an effective way to validate the results obtained using one method by undertaking the same experiment using another method and reach to the same or close results (Thurmond 2001).

Kennedy (2009) suggested that bias can be reduced using triangulation through:

i. **Measurement bias:** caused by the circumstances and setup involved with the data collection, e.g., traffic condition. Triangulation can minimise the measurement bias by recollecting the data under different circumstances and compare between them, e.g. different traffic conditions;
ii. **Sampling bias:** concerned with the quantity of the collected data and whether it is sufficient for the research scope. Triangulation can help researcher collect and generate data to ensure that sufficient sampling of data to represent the research problem is achieved; and
iii. **Procedural bias**: focuses on the quality of the collected results. The method can have a direct effect on the reliability of results, e.g., uncertainty and errors in simulation. Triangulation can reduce procedural errors in the obtained results by combining methods of different procedures, e.g., smoothing and replication of simulation results.

This research adapts the mixed-methods or triangulation choice. Quantitative methods count for the majority of the work, while qualitative aspects of the research were taken into consideration to reduce bias in the collected data and obtained results and increase the validation efficiency.

The quantitative methodology was implemented in the following order:

i. identify variables to be investigated;
ii. define research methods to be used to analyse the variables; and
iii. carry out experiments on the variables and interpret the results.

This research applied, in addition to the quantitative methods, qualitative evaluations to deal with the limitations of quantitative research mentioned in 4.2.2.1 and reduce the bias related to the qualitative aspects and nature of the variables and the author interpretations.

The research addressed the bias aspects during the data collection, pre-processing and analysis stages of the study. Furthermore from the validation perspective, the developed autonomous control method was validated using both a flow line model developed by other researcher (Scholz-Reiter et al. 2005) and a real-world case study.

4.2.3 Methodological Framework

4.2.3.1 Data Mining Framework

The Data Mining Framework used in the development of the evaluative model is based on a developed search approach referred to as the '*Degree of Freedom (DOF)*'. The DOF approach was developed to standardise the implementation process of data mining and helps to choose the suitable methods which can reduce bias for each research step. The approach investigated the degrees of freedom imposed at each step of the evaluative model development and decided the data mining methods to be applied. The search steps based on the DOF approach were broken down into three phases:

i. Phase I – Data Pre-processing;
ii. Phase II – Feature Selection; and
iii. Phase III – Model Building.

Implementation of the DOF approach within these phases is elaborated in the corresponding research methods and steps, i.e., Section 4.5.2, 4.5.3.1 and 4.5.3.2 respectively.

The framework is validated using the real-world case study as described in Section 4.5.4.1.

4.2.3.2 Autonomous Control Framework

The developed empirical formula evaluative model was used here as the foundation for building an autonomous control method for flexible flow lines. The formula-based autonomous control used the built-in formulated relationship to evaluative the impact of the decision step on the throughput rate and determine the optimal decision.

The developed autonomous control method starts by estimating the throughput rate of the variability scenario according to the decision variables and the constraints imposed on the system. Then the decision step is undertaken to optimise the throughput rate of each product type, while considering the system stability for each decision variable and that there is no overlap between executed decisions for multiple product types, e.g., two products sent to the same process at the same instance. Details of the research methods and steps involved are given in Section 4.5.3.3.

The framework is tested, in a simulation environment, using representative variability scenarios of processing times and arrival rates for a 3x3 flow line model developed by Scholz-Reiter et al. (2005) and a real-world case study as described in Section 4.5.4.2.

4.3 Research Design

Variables can be independent, dependent or control variables. The research outcome, in its core, is to accurately estimate and formulate the degree of dependence and correlation between the investigated variables and use them to control the variability in human-dependent serial flow lines.

From the model building perspective, the three main process-based parameters that represent the flow line, μ_i, c_i and N, remain constant for synchronous flow lines, hence, use of these terms as predictors is appropriate. However, use of these variable parameters as predictors for asynchronous flow lines lacks the potential of being generic representation of a flow line. Furthermore, since it is a case-by-case form, it will require an enormous number of scenarios to represent long flow lines. Finally, the use of discrete processing times and locations as predictors is not appropriate, as discussed in Section 1.2 and 2.6.2. Hence, a new set of generic parameters are investigated in this study to represent the variability of both μ_i, c_i within the flow line with minimal number of variables. The degree of dependency between these variables and their relationship to the dependent variable TR is part of the investigation.

As shown earlier in Section 2.6.1, the maximum processing time plays an important role in the throughput rate of non-exponential flow lines. In fact, the bottleneck, i.e., the process with the maximum actual processing time, governs the throughput rate for deterministic processing times (First term in Equation 1). However for non-deterministic flow lines (Second term in Equation 1), such as the case in this research, the process with the maximum mean and maximum actual processing time do not always match. The bottleneck can constantly move based on the mean processing times along the flow line and the average coefficient of variation, i.e., when the actual processing time of a process exceeds the maximum mean processing time. Furthermore, the use of processing times of each process, in addition to what was explained earlier in Section 2.6.1, will require an enormous number of independent variables to represent long flow lines. Hence, additional generic data mining-compatible parameters, i.e., the minimum, average and coefficient of variation of mean processing times within the flow line, can explain the discrepancy due to the potential movement of the bottleneck for asynchronous non-exponential flow lines. They essentially represent the proximity of the processing times in respect to the flow line, hence, the potential movement of the bottleneck.

Furthermore, researchers did not investigate the effect of the location of the process with the maximum mean processing time within the flow line, i.e., the ratio between the process with the maximum mean processing time and the length of the flow line.

Hence, these four parameters were added along with the ones from Li and Meerkov (2009) formula (Equation 2.8) as follows:

i. Minimum Mean Processing Time within Flow Line $(\mu_{\min})$:

$$\mu_{\min} = \min_i \mu_i \tag{4.1}$$

ii. Average Mean Processing Time within Flow Line (μ):

$$\mu = \frac{1}{N}\sum_{i=1}^{N}\mu_i \tag{4.2}$$

iii. Maximum Mean Processing Time within Flow Line $(\mu_{\max})$:

$$\mu_{\max} = \max_i \mu_i \tag{4.3}$$

iv. Coefficient of Variation of Mean Processing Times within Flow Line (c):

$$c = \frac{1}{\mu}\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mu_i - \mu)^2} \tag{4.4}$$

v. Average Coefficient of Variation (c_{av}):

$$c_{av} = \frac{1}{2N}\sum_{i=1}^{N} c_i \tag{4.5}$$

vi. Location Ratio of the Process with Maximum Mean Processing Time (l):

$$l = {}^{i}\!/_{N} \text{ such that } \mu_i = \mu_{\max} \tag{4.6}$$

vii. Length (N)

Furthermore, the investigation includes the direct and multiplicative inverse of linear and nonlinear terms of each variability parameter, i.e., variable. The general criteria for election of parameter terms as model predictors were set as:

i. Only terms with highly strong relationships to *TR* were considered, i.e., correlation coefficient equals or higher than 0.8 (Baird 2010); and
ii. Relationship is considered insignificant and the predictor terms excluded if the *p*-value is higher than 0.1 with the following levels (Mendenhall et al. 2012) used for evaluation of the significance:
    a. Highly significant: *p*-value is less than 0.01;
    b. Statistically significant: *p*-value is higher than 0.01 but less than 0.05;
    c. Possibly significant: *p*-value falls between 0.05 and 0.1; and
    d. Insignificant: *p*-value is higher than 0.1.

The regression covariates in stepwise regression were also elected, or de-elected, using the same criteria.

4.4 Tools

4.4.1 Discrete Event Simulation Package

Modern simulation modelling software provides high flexibility to represent complex flow lines and a visualisation platform (Haik and Aomar 2006). It is also a helpful tool for verification and comparison of analytical modelling results. If the process was considered as a simulation modelling element, the changes to it are not continuous in respect of time, instead go through step changes such as receive product, process, dispatch, etc. and different states, e.g., busy, blocked, stopped, etc. Discrete Event Simulation (DES) modelling is "*based on a discrete internal representation of model variables*" (Haik and Aomar 2006). Therefore, DES model is a preferable simulation modelling option of operations within flow lines (Papadopoulos et al. 2009).

As discussed in Section 1.2 and 2.6.2, the theoretical route to obtain the throughput rate for non-exponential flow lines is currently not possible and real-world observations are constrained to specific variability cases. Hence, simulation is an appropriate tool to determine the throughput rate for a wide range of generic representative variability scenarios of non-exponential flow lines. Furthermore, simulation has been the standard benchmark for performance evaluation of empirical and theoretical evaluative models of flow lines, e.g., Blumenfeld (1990), Li and Meerkov (2009) and Wang et al. (2014).

The main advantages of simulation modelling are as follows:

i. **Repeatability:** Experiment using simulation models with similar parameters will lead to the same results (Nehmzow 2009);
ii. **Reliability:** Simulation models developed significantly and the accuracy of obtained results becomes high and close to actual systems; despite the fact that all models are inherently not similar to actual system (Box and Draper 1987);
iii. **Design Support:** Allow different scenarios of operations to be investigated during the design stage before real-world implementation (Nehmzow 2009);
iv. **Adaptability to any Operations:** Flexible enough to model operations of service or production-based sectors (Haik and Aomar 2006); and
v. **Virtual Analysis:** Provide virtual environment to experiment the operations without the incurring cost of real implementation (Nehmzow 2009).

Simulation modelling has also drawbacks such as (Haik and Aomar 2006):

i. The capital cost of acquiring the simulation software and operational cost in terms of man-hours and overheads;
ii. Development time to build simulation model can be long;
iii. Extensive data collection and validation are usually required to ensure the simulation model reflects the real-world scenario.

The advantages of simulation model outweigh the disadvantage from the research and the InnovateUK project perspectives. Therefore, a DES package, Simul8, was used in this research to model the variability within a generic representation of human-dependent serial flow lines, Scholz-Reiter (2005) 3x3 flow line model and the real-world case study. It was then used to obtain the throughput rate for each individual variability scenario. Simul8 was also used as a comparison platform for the autonomous control and simulation-based optimisation, i.e., OptQuest, techniques.

4.4.2 MATLAB

MATLAB was mainly used in this research for model building of the evaluative regression machine learning models for synchronous and asynchronous human-dependent serial flow lines and to compare these models against classification machine learning models and existing formulas from the literature. Statistics and Machine Learning and Neural Network toolboxes of MATLAB provide a wide range of classification and regression machine learning algorithms. They were used in this research to build and examine the following supervised machine learning models:

i. **Regression:**
   a. Multiple Linear Regression:
      - *Robust Regression;*
      - *Stepwise Regression*; *and*
      - *Generalised Regularisation Linear Models; and*

ii. **Classification:**
   a. Feedforward Neural Network:
      - *Levenberg-Marquardt backpropagation; and*
   b. Decision Tree:
      - *Bootstrap Aggregating; and*
      - *Boosting.*

4.5 Research Methods and Steps

This section describes the generic methods and steps related to the research. The research steps related to the case study used for validation is described in Chapter 5. Figure 4.3 shows the research steps and the type of methods used in each.

4.5.1 Data Generation

Synthetic data were generated to ensure that the developed methods can be applied outside the specific case study of this research. Complexity was introduced gradually to the data set to cover a wide range of variability scenarios that can occur in a synchronous and asynchronous human-dependent serial flow line. Synthetic discrete data were used throughout the development phases while the actual continuous data were applied to the validation stage.

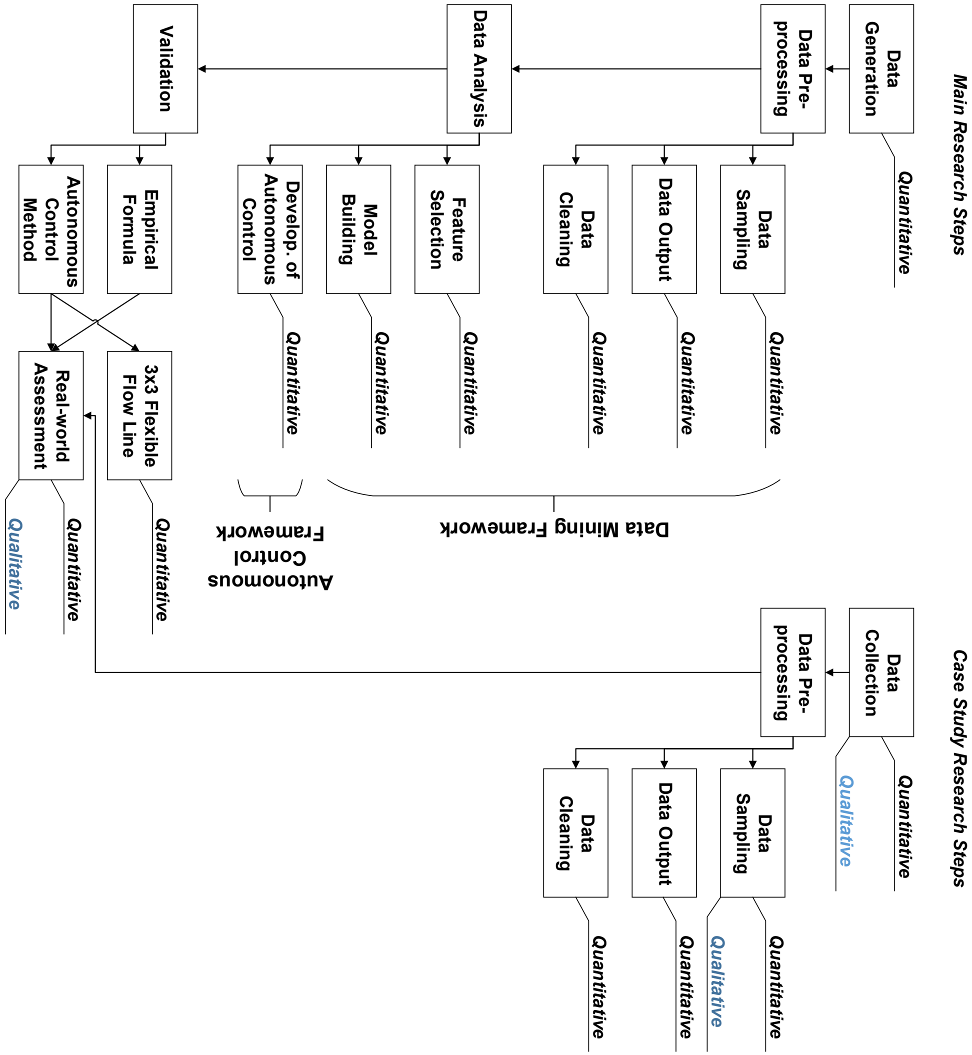


Figure 4.3: Research Steps

4.5.2 Data Pre-processing

Sampling of data represents the first challenge for data mining modelling of any real-world system especially when large data and inter-dependent variables are involved (Yang and Wu 2006). Accuracy of the model tends to be directly related to the sampling size used. Furthermore, large sampling size helps the predictive model to be more representative of the actual system and reduces the chances of overfitting (Weiss 1998). The drawback is that the algorithm used to build the model can become computationally extensive and the model itself tends to be more complex with insignificant improvement to accuracy (Oates and Jansen 1998). Therefore, the data sets were chosen to be large enough to represent real-world flow lines with arbitrary scenarios of variability.

The process of sampling divides the data into three classes; Class I (the training set), Class II (the test set) and Class III (the validation set). Class I and II are used during model building to discover and generalise the patterns to the population while Class III is applied to another set of data for validation (Hastie et al. 2009).

4.5.2.1 Data Sampling

4.5.2.1.1 Synchronous Non-exponential Serial Flow Line

Data sets were generated to represent the intra-variability of synchronous non-exponential serial flow lines. The stochastic nature of such flow lines exists within the individual process only and no inter-process variability exists across the flow line. Hence, the data sets were created to represent the intra-variability of processing times, μ and c of an individual process P and the length N of the synchronous flow line.

Four discrete data sets were used for training and testing of the evaluative models of synchronous flow lines (Class I and II).

The data sets $I/II-S-1$ to $I/II-S-3$ are generated to fully represent the processing time variability up to a scale of $10, 1 \le \mu \le 10$, from a short $N=2$ to a relatively long flow line, $N=21$ with the three parameters μ, c and N. The first and second data sets ($I/II-S-1$ and $I/II-S-2$) included all scenarios for a single and two parameters respectively and selective level for the remaining parameter(s) ($I/II-S-1$ is shown in Appendix B (P. A-4)).

Full factorial Design of Experiments (DOE) was applied to generate all variability scenarios for the data set $I/II-S-3$. Finally, the scaled-up data set, $I/II-S-4$ was generated using full factorial with higher mean processing time variability, $1 \le \mu \le 10$, and longer flow lines $N \in \{23,30\}$.

The variability range of generated data sets is:

$I/II-S-1$, $I/II-S-3$: $\mu \in \{1,2,...,10\}, N \in \{1,2,...,21\}$,

$I/II-S-2$: $\mu \in \{1,3,5,7,9\}, N \in \{2,3,4,6,8,10,12,14,16,18,20,21\}$, and

$I/II-S-4$: $\mu \in \{1,2,...,15\}, N \in \{23,30\}$,

and $c \in \{0,0.05,0.15,0.45,1\}$ for data set $I/II-S-2$ and

$c \in \{0,0.01,0.025,0.05,0.075,0.1,0.25,0.5,0.75,1\}$ for data sets $I/II-S-1$, $I/II-S-3$ and $I/II-S-4$.

4.5.2.1.2 Asynchronous Non-exponential Serial Flow Line

Eight discrete data sets were sampled. Representative data sets were defined for training and testing of the intra- and inter-variability of processing times P_i and length N within asynchronous non-exponential flow lines. The first two data sets ($I/II-A-1$ and $I/II-A-2$) were chosen to fully represent the processing time variability up to a scale of 10, $1 \le \mu_i \le 10$, for a relatively small flow line, $N < 5$. Data set $I/II-A-1$ is for flow lines with lengths of one and two processes while three and four processes are covered in the data set $I/II-A-2$.

For flow lines with one to four processes, full factorial DOE was used to generate all scenarios in the data set, where mean processing times varies between 1-10 time units. This was considered since full coverage of this domain of flow lines should be applicable for scaling-up to cover flow lines with arbitrary μ_i, c_i and N variations.

For longer flow lines, μ_i was selected randomly and equiprobably for the second six data sets as follows:

$I/II-A-3$: $\mu_i \in \{1,2,...,10\}, N \in \{1,2,3,4,5\}$,

$I/II-A-4$: $\mu_i \in \{1,2,...,60\}, N \in \{1,2,3,4,5\}$,

$I/II-A-5$: $\mu_i \in \{1,2,...,100\}, N \in \{1,2,3,4,5\}$,

$I/II-A-6$: $\mu_i \in \{1,2,...,60\}, N \in \{1,2,...,30\}$,

$I/II-A-7: \mu_i \in \{1,2,...,100\}, N \in \{1,2,...,30\}$, and

$I/II-A-8: \mu_i \in \{1,2,...,500\}, N \in \{1,2,...,30\}$,

and $c_{av} \in \{0,0.01,0.025,0.05,0.075,0.1,0.25,0.5,0.75,1\}$ for all data sets, i.e., $I/II-A-1$ to $I/II-A-8$.

4.5.2.2 Data Output – Simulated Throughput Rate

In this step, the output of the synthetic data was obtained using simulation. Simulation models were developed based on the variability scenarios represented by the synthetic discrete data sets ($I/II-S-1$ to $I/II-S-4$ and $I/II-A-1$ to $I/II-A-8$) using Simul8 simulation package.

The simulation model was programmed to loop through the scenarios and determine the steady state simulated throughput rate for each variability scenario.

For a given simulation model, steady state analysis can be based on:

i. Known Initial Conditions: i.e., the modelling elements are set with initial conditions that satisfy the steady state. For example, the queues and machines will have some work items and uncompleted jobs respectively. The model here replicates the conditions at the start of a normal day in the actual flow line; and
ii. Unknown Initial Conditions: either initial conditions are not known or are difficult to predict (Hoad et al. 2008):
    a. **Warm up:** Run model for a long period of time until the performance measures reach the saturation state and delete this period (warm-up period) from collected results;
    b. **MLE:** Use the Maximum Likelihood Estimator (MLE) method to determine the steady state performance measures even when the system might still be in the transient state (Sheth-Voss et al. 2005); and
    c. **Infinity:** Set the simulation time to be long enough $t_{sim} \rightarrow \infty$ to reduce the effect of transient period.

Since warm-up period changes from one experiment to another based on the variability of the system and incurred cost of running the model for a long period of time is not an issue, the '*Infinity*' method was chosen in this study to obtain the steady state throughput rate.

4.5.2.3 Data Cleaning

This step is concerned with reducing the noise of the output data, i.e., simulated throughput rate, to prepare the data for the data mining analysis. Traditionally, replication of simulation runs is used to increase the confidence interval (CI) of a certain performance measure of interest (Law and McComas 1990, Robinson 2004, Banks et al. 2005, Law 2007).

Two data cleaning methods were examined; Robinson (2004) confidence interval method to determine the number of simulation replications required for the desired CI and the second is smoothing of the output data. The investigated DOFs during this phase are (Figure 4.4):

i. **DOF-I-1 Simulation Replication:** repetition of the simulation runs to obtain an average value of each output value with higher confidence interval; and
ii. **DOF-I-2 Smoothing:** application of a smoothing method to the complete simulation outputs in a single data set to reduce the noise and detect outliers

In DOF-I-1, the models were configured to import the data set, loop through the scenarios within each data set and obtain the steady state throughput rate with the number of runs that satisfy the conditions of 0 and 95% confidence intervals for each scenario, i.e., single and multiple runs respectively.

The optimal number of iterations to give a CI of 95% for each experiment was determined using the method proposed by Robinson (2004). Afterwards, a comparison between the steady state throughput rate with 0% and 95% CI, i.e., single and multiple runs respectively, was carried out using the data set $I/II-A-1$ to obtain the main source of error. The Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) of the steady state simulated throughput rate were determined and the correlation of this error to the parameters of the flow lines was investigated.

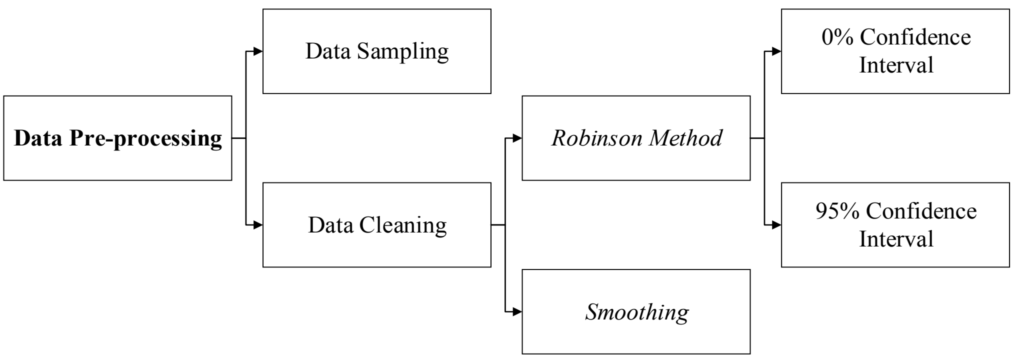


Figure 4.4: DOF of Phase I

In addition to the Robinson (2004) replication method, DOF-I-2, i.e., smoothing, was applied to the simulation output to reduce noise for comparison reasons. The smoothing method is a widely acceptable approach for data cleaning (Jeffery et al. 2006).

Smoothing was applied to the data sets $I/II-A-1$ and $I/II-A-2$ only since these two sets were arranged in a coherent order from the least to the highest variability scenario, hence, smoothing is applicable. Other sets, apart from $I/II-S-3$ for synchronous flow lines which has considerably smaller number of scenarios, are randomly selected, i.e., the neighboring data points within the defined span are not correlated since the *x*-axis here is the experiment number which moves from one random scenario to another. Hence, smoothing of the output data for these data sets is not feasible. The following smoothing techniques were applied using MATLAB:

i. Moving Average;
ii. Savitzky-Golay Filter.
iii. Local Regression:
    a. 1st degree polynomial; and
    b. 2nd degree polynomial model; and
iv. Robust Regression:
    a. 1st degree polynomial; and
    b. 2nd degree polynomial model.

This step investigates if the error generated by the simulation software can be mitigated by smoothing of the simulation output data for ordered data, though this imposes a major limitation on this data cleaning technique regardless of the performance of the method. Results are reported in Appendix C (P. A-6).

4.5.3 Data Analysis

After the synthetic data were prepared in the data-preprocessing stage, analysis was carried out on the data according to the frameworks described in section 4.2.3 to build the relationship between variables into a formula-based evaluative model and use the model to develop an autonomous control method.

4.5.3.1 Feature Selection

This phase provides a new representation of asynchronous non-exponential serial flow lines using selected linear and nonlinear terms of line-based parameters based on their impact on the throughput rate TR. The investigation includes impact and stability analysis of each prediction line-based parameter on TR using statistical analysis.

The relationship between each parameter, including its linear and nonlinear terms, and TR was studied to determine whether or not a relationship exists and to what extent this relationship is significant. Removal of line-based parameter terms with no effect on the throughput rate as predictors from the model is an acceptable approach to improve prediction accuracy since they do not represent features of the modelled dependent variable, i.e., TR. Furthermore, the multiplicative inverse of linear and nonlinear terms of each parameter was included in the analysis since TR can be directly related to one or both of them, i.e., exists in both the nominator and denominator of the formula.

It is worth mentioning that the study of the effect of each parameter on TR individually is not always possible since the change in some parameters automatically disturbs the others due to the inter-dependency between these variables, e.g., $\mu_{\max}$ and μ. Therefore, for each parameter, a data set (Class IV) was created with the smallest number of factorial changes for each sub-set.

Statistical analysis was carried out to determine the strength and significance of the relationship between parameters and TR. Correlation analysis was applied to examine the strength of the relationship. However, to determine the significance of this relationship, Analysis of Variance (ANOVA) was performed on the data set; *f*- and *p*-value of regression coefficients and *f*-value of regression model were examined to determine if the parameter term is statistically significant. Finally, best sub-set regression was applied to verify the results and determine if a parameter term can be excluded from the model building stage.

It is worth noting here that the variability within synchronous flow lines is generated inside the process only, i.e., intra-variability, hence, the process-based and line-based parameters are the same. Therefore, the selected line-based parameter terms were then translated to process-based parameter terms for the special case of synchronous flow lines to be used as predictors in the model building phase.

This phase includes the following DOFs (Figure 4.5):

i. **DOF-II-1 Representation Type:** the parameters used to represent the variability within each individual process or the flow line; and

ii. **DOF-II-2 Parameters Terms:** the data mining models are trained with the direct linear term of the parameters or the direct and multiplicative inverse of the linear and nonlinear terms.

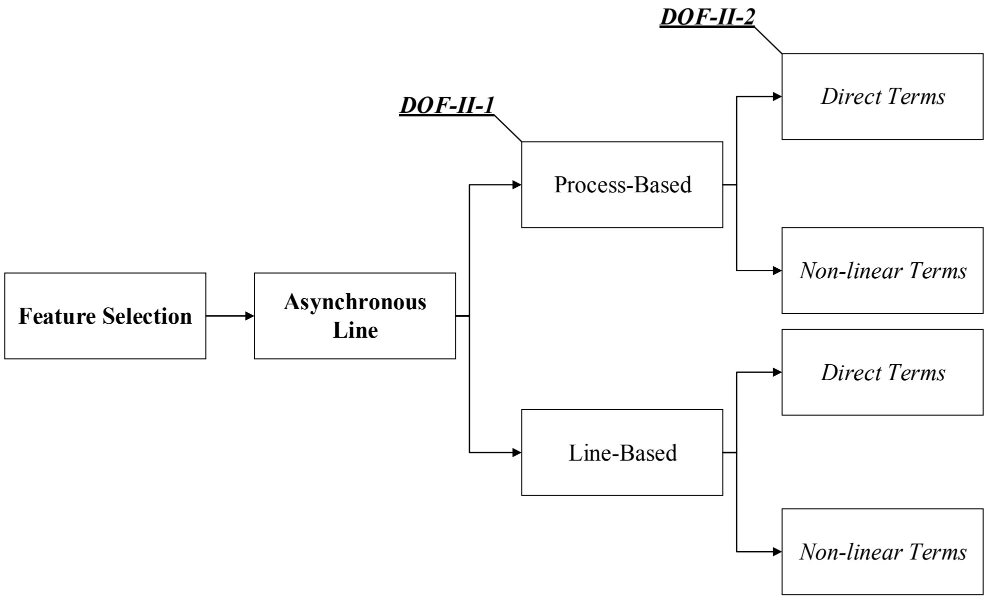


Figure 4.5: DOF of Phase II

4.5.3.2 Model Building

A new set of line-based parameters to represent the variability within the stochastic asynchronous flow lines were defined. The line-based parameter terms were then shortlisted in Phase II based on their influence on the throughput rate for asynchronous flow line and redefined in terms of process-based parameter terms for synchronous flow lines. In this phase, the selected predictors with the highest impact on *TR* were used to formulate the relationship between the variability in the flow line and the throughput rate. Data mining techniques were used to generate and validate the formula-based evaluative model.

The derivation of the empirical formulas includes the following main steps:

i. Development of a MATLAB program to analyse the performance of a set of data mining models based on supervised machine learning techniques and current formulas from the literature; and

ii. Determination of the optimal standalone empirical formulas of the throughput rate of synchronous and asynchronous flow lines as a function of the process-based and line-based parameters, respectively, with the highest impact on the throughput rate.

Two evaluative models were built during this phase for synchronous and asynchronous flow lines. The two models used different cross validation partitioning criteria and DOE techniques to suit the nature of the data sets and for comparison reasons. The DOFs during this phase can be classified in general into the following categories (Figure 4.6):

i. **DOF-III-1 Supporting Predictors:** selection of the supporting predictors to be included in the training of the data mining model;
ii. **DOF-III-2 Cross Validation Partitioning:** data sets assignment to the training and test sets; and
iii. **DOF-III-3 Modelling Method:** supervised machine learning regression to build the formula-based evaluative model of the throughput rate:
    a. *DOF-III-3a Stepwise Regression – Model Type:* The following forms of regression are considered:
        - *Interaction: covariates can be a single or multiplication of two linear (1st degree polynomial) predictor term(s);*
        - *Purequadratic: linear and squared (2nd degree polynomial) terms are included in this model;*
        - *Quadratic: comprises linear and squared (2nd degree polynomial) predictor terms and multiplications of two linear terms;*
        - *Polynomial: includes multiplication of linear predictor terms up to the 6th polynomial degree;*
    b. *DOF-III-3b Stepwise Regression – Bounded and Unbounded Steps:* In unbounded steps, the model starts with the full regression model and removes covariates from it or adds ones from lower regression forms, e.g., interaction for purequadratic model. In bounded steps, however, the model is built in forward iteration inclusively from the specified regression model;
    c. *DOF-III-3c Robust Regression:* the use of different fitting techniques for the least squares; and

d. *DOF-III-3d Regularisation Algorithms:* three algorithms for regularisation of the least squares.

The Ordinary Least Squares (OLS) regression was excluded from the set of regression machine learning models since it lacks robustness or adjustments to the squares error to restrict the covariates to the significant variables and variable relationships. Furthermore, two types of stepwise regression models were also not considered, namely '*Constant*' and '*Linear*' models, since predictor terms were chosen based on significance and strength of their relationship to the throughput rate as described in Phase II '*Feature Selection*', hence, removal of all of them is not a DOF. As for '*Linear*' regression model, it was substituted by Regularisation Algorithms as they perform the same function.

4.5.3.2.1 Derivation of Empirical Formula

- **Synchronous Flow Line**

A MATLAB program was built to examine the evaluative models. The program was configured to read the data, build the supervised machine learning models and perform statistical analysis on the results to evaluate the model performance.

The model building process as shown in Appendix D (P. A-8) (Figure D.1) runs through the different degrees of freedom in model building. It starts by importing all the data sets $\mathbf{D}$ to the MATLAB program. The individual data set $I/II-S-1$ to $I/II-S-4$ are segregated to $\mathbf{D}_x, x \in \{1,2,3,4\}$. The process-based parameter terms as selected in Phase II, are then identified as the model predictors $p_y, y \in \{1,2,3,4,5\}$ such as the predictors set $\mathbf{P} \subset \mathbf{D}$. A counter of the training set number $w, w \in \{1,2,3,4\}$ is then started. The training and test sets, $\mathbf{S}$ and $\mathbf{T}$ respectively, for the current experiment are then defined before starting to run through the regression and classification machine learning models. The models were trained and tested using the cross validation technique with step partitioning of the training and test sets, i.e., iterative selection from the data set $\mathbf{D}$.

Figure 4.6: DOF of Phase III

The data sets used for training were excluded from the main test set $\mathbf{T}_o$ but included in the supporting test set $\mathbf{T}_u$ to examine the models for overfitting. The following machine learning methods were used for data mining and model building:

i. **Regression:**
   a. Multiple Linear Regression;
      - *Robust (M-estimators) Regression:*
        - Hinch and Talwar (1975) method;
        - Andrews (1974) method;
        - Tukey's Bisquare (1960) method;
        - Holland and Welsch (1977) method;
        - Cauchy M-estimators by Moore (1977);
        - Huber (1981) method;
        - Fair method by Rey (1983); and
        - Logistic regression;
      - *Stepwise Regression:*
        - Interaction;
        - Purequadratic;
        - Quadratic; and
        - Polynomial:
          - *3rd to 6th degree; and*
      - *Regularisation Algorithms:*
        - Lasso;
        - Ridge Regression; and
        - Elastic Nets; and

ii. **Classification:**
   a. Feedforward Neural Network:
      - Levenberg-Marquardt Backpropagation; and
   b. Decision Tree:
      - Bootstrap Aggregating; and
      - Boosting.

Also, the following formula from the literature was added for comparison purposes:

iii. Blumenfeld (1990).

The Levenberg-Marquardt Backpropagation Neural Network, used in this research, had one input layer with number of neurons equals to the number of predictors, 20 hidden layers and one output layer with a single node. Performance of the Neural Network is determined using the Mean Absolute Errors. As for the Decision Trees, Bootstrap Aggregating and Boosting Decision Trees were configured to have 100 and 1000 ensemble learning cycles respectively.

The program runs through each method m and calculates the MAPE for each test set $\mathbf{T}_a$:

$$e_{ma} = \frac{100}{q_{T_a}} \sum \left| \frac{\mathbf{TR}_a^{sim} - \mathbf{TR}_{ma}^{pred}}{\mathbf{TR}_a^{sim}} \right| \tag{4.7}$$

where

e_{ma} is the absolute %error of method m for the data set a within the test set $\mathbf{T}$;

q_{T_a} is the number of variability scenarios within the test set $\mathbf{T}_a$;

$\mathbf{TR}_a^{sim}$ is the simulated TR of the scenarios within the test set $\mathbf{T}_a$; and

$\mathbf{TR}_{ma}^{pred}$ is the predicted TR using method m of the scenarios within the test set $\mathbf{T}_a$.

The results for each training experiment $\mathbf{R}_w$ are then collated to the set $\mathbf{E}$. The mean and standard deviation of the errors of each method m and data set a within the test set $\mathbf{T}$, i.e., μ_e and c_e, are calculated for the set $\mathbf{E}$ to determine which method outperforms the others for the particular training set $\mathbf{R}_w$.

The error percentages are then rounded to the nearest hundredth and $\mu Score$, $cScore$ of each method m within the set $\mathbf{E}$ of the training experiment $\mathbf{R}_w$ are determined and compared to obtain the method(s) that performed the best with minimal errors according to the scoring criteria shown in Table 4.1.

Subsequently, the program re-runs through the same steps but with another data set $\mathbf{D}_x$ used for the training of models. After the program runs through all the possible data sets, the data set $\mathbf{S}_x$ of the training set $\mathbf{R}_w$ which includes the best performing method for all $\mathbf{D}_x, x \in \{1,2,3,4\}$ is elected and saved.

Finally, the model parameters, i.e., the covariates $\mathbf{V}$ and regression coefficients $\boldsymbol{\beta}$, of the best performing method are extracted and performance plots are generated. The following plots are generated to visualise the goodness of fit:

i. predicted TR vs. actual (simulated) TR;
ii. predicted TR vs. residuals in TR;
iii. correlation between residuals in predicted TR;
iv. histogram of errors; and
v. normal probability plots of residuals in predicted TR.

Furthermore, the contribution of the predictors to the changes in TR were also obtained by extracting the '*major effects of factors*' for the best performing model.

Table 4.1: Scoring Criteria for *μScore* and *cScore*

Rounded μ_e, c_e *to Hundredth*	$\mu Score$, $cScore$
>=100%	**0**
20-99%	**1**
10-19%	**2**
9%	**3**
8%	**4**
7%	**5**
6%	**6**
5%	**7**
4%	**9**
3%	**11**
2%	**13**
1%	**15**
0%	**20**

- **Asynchronous Flow Line**

The eight data sets ($I/II-A-1$ and $I/II-A-8$) were modified to include the shortlisted main predictors $\mu_{\max}{}^{-1}$, c_{av}, $e^{c_{av}}$ and N, and also the supporting predictors $e^{\mu_{\min}{}^{-1}}$, μ, μ^{-1}, $\log\mu$, $e^{\mu^{-1}}$, c, c^{-1} and $\log c$.

In terms of the TR, it was obtained two times using the simulation model, once at the actual c_{av} and the other with c_{av} equals to 1, for each variability scenario within the eight data sets. The simulated TR at $c_{av}=1$ was obtained to be used during the modelling stage to feed the empirical formula developed by Li and Meerkov (2009) which is one of the comparison models.

To carry out the model building process, an extended version of the MATLAB program for the synchronous flow line was developed (Figure D.2). The models used with the asynchronous flow line case for data mining and model building are the same as for the synchronous case listed in Section 4.5.3.2.1.

In addition, the following formulas were used for comparison purposes:

iii. Deterministic throughput rate TR^d that satisfies the condition $c_i = 0, i = 1,2,....,N$, i.e., $1/\mu_{\max}$; and

iv. Li and Meerkov (2009).

The program runs in the same manner as for the synchronous flow line case but with more DOFs. For instance, there is an additional DOF related to the ability to use multiple data sets $\mathbf{S}_n, n \in \{1,2,...,8\}$ for training of the models. The data set here is defined as $\mathbf{D}_x, x \in \{1,2,..., X\}$, where the variable X determines the number of data sets within $\mathbf{D}_x$ that can be used for training when n increases, after exclusion of the best performing data set $\mathbf{S}_n$ from the data set pool $\mathbf{D}_x$, $X = 8-n+1$. This DOF requires that the program checks that the addition of a new data set is feasible by checking that the $\max \boldsymbol{\mu}\mathbf{Score}$ at elected $\mathbf{S}_n$ is greater than at elected $\mathbf{S}_{n-1}$, otherwise the program will stop training and compile elected $\mathbf{S}_{n-1}$ into the optimal $\mathbf{R}_w$.

Another DOF is the existence of supporting predictors $p_j, j \in \{1,2,3,...,8\}$. Iterative inclusion of the supporting predictors to reach the optimal set of supporting predictor(s) $\mathbf{S}_j$ is also applicable using the same procedure as described for the training data sets. The supporting predictors' procedure is commenced after the optimal training data set(s) $\mathbf{S}_n$ are reached.

The elected set(s) $\mathbf{S}_j$ including the elected data set(s) used for training $\mathbf{S}_n$ are compiled into the optimal training set $\mathbf{R}_w$ and the covariates and regression coefficients of the best performing method within this set are extracted and results are plotted.

4.5.3.3 Development of Formula-Based Autonomous Control Method

In a simulation-based flow line modelling, optimisation and autonomous control are driven by the simulated local information of performance measures at the process level or system information of the entire flow line. The objective function for optimisation in this case is the simulated local- or system-level performance measures. This research presents a formula-based autonomous control method that can suggest the optimum solution independent of the simulation model.

The formula-based autonomous control logic for routing decisions works by;

i. setting decision steps equal to the number of processing steps;
ii. evaluates the throughput rate for possible decisions;
iii. chooses job sequence for each product that produces highest *TR* and minimal queue times, i.e., no overlap between routes for different products; and
iv. adjusts the routing decision based on sudden changes within the flow line, e.g., breakdown.

For decisions on variability factors, the logic is as follows;

i. chooses one random decision variable;
ii. gets the optimal setting for it, where throughput rate starts to stabilise, i.e., increases of less than 1% between sequential steps;
iii. adds the next variability factor;
iv. gets the corresponding optimal settings at the verge of stability of the throughput rate and so on.

Figure 4.7 shows a visual representation of the second general case.

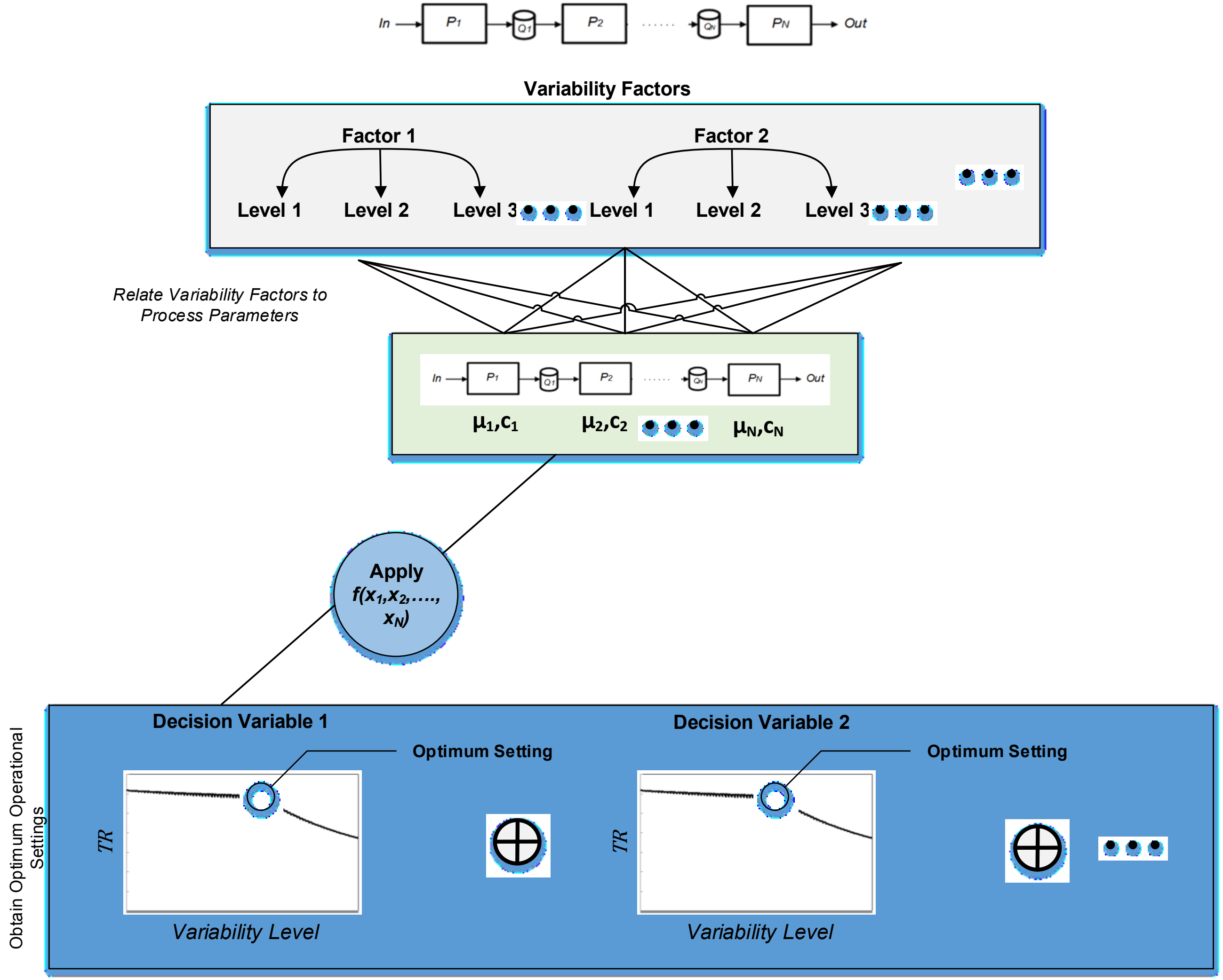


Figure 4.7: Formula-Based Autonomous Control Logic

4.5.4 Validation

4.5.4.1 Empirical Formula

In terms of testing the regression empirical formula, the model was compared in Phase III with other classification machine learning models and existing formulas from the literature. However, these models were trained and tested using synthetic discrete data. In this section, the formula was validated with continuous actual data from a real-world industrial case study. This step is concerned with testing the prediction accuracy of the empirical formula and its validity to real-world scenarios.

The simulated throughput rate was compared against the calculated throughput rate using the empirical formula and the average, minimum and maximum mean absolute percentage errors, MAPE, MINAPE and MAXAPE, were determined. Correlation analysis was then applied to investigate the variability factor(s) and the predictor term(s) that contribute the most to the residuals.

4.5.4.2 Autonomous Control Method

The performance of the developed formula-based autonomous control method was compared to existing simulation-based optimisation and autonomous control methods within a DES modelling environment to study its strengths and limitations.

The formula-based autonomous control logic can be applied for a routing decision in a flexible manufacturing system. For instance, a 3x3 flexible flow line with three different products was chosen for the validation.

4.6 Summary

This chapter covered the methodology and generic research data generation, pre-processing, and analysis and validation methods and steps. It also gave details regarding the two methodological frameworks. Mixed-methods were chosen as they give the best of both worlds; apply to a quantitative research, and multi methods including qualitative methods can be used in a single step. This enabled development of a search approach that governed the Data Mining framework which was extended to the Autonomous Control Framework. The research strategy combines this generic research and experiments with a real-world case study.

The next chapter describes the specific research methods and steps related to the real-world case study. Chapter 6 and 7 present the results related to research methods and steps within the two methodological frameworks respectively.

5 PILOT STUDY

5.1 Overview

This chapter covers the steps and methods of the real-world case study used in this research to represent an industrial example of flexible flow lines.

5.2 Case Study

This research is part of a collaborative research project between De Montfort University and industrial partners funded by Innovate UK (InnovateUK Grant No. 18834-132285 '*Development of an innovative Autonomous Model Development Tool (AMDT) for boosting manufacturing process competencies*'). The project aim is to develop a new 'Autonomous Process Model Development APMD' approach capable of precisely identifying where and when business-orientated process innovation is necessary, what aspects of processes need innovating and how successful are the new process competences. This is achieved through:

- Process mapping and identification of process variability of specific case studies from the industrial partners where business-orientated process innovation is required.
- Analysis of the causal relationships between the controllable and uncontrollable variability factors and performance measures of interest for the industrial partner.
- Manipulate the controllable factor(s), i.e., decision variables, in order to reach the optimal performance measures of the system using the defined causal relationships.

Figure 5.1 shows where the research outcomes lie within the InnovateUK-funded project objectives. The case study reported here is done collaboratively with Costain Group plc. Costain is a British engineering company, and the investigation is focused on a major motorway development project.

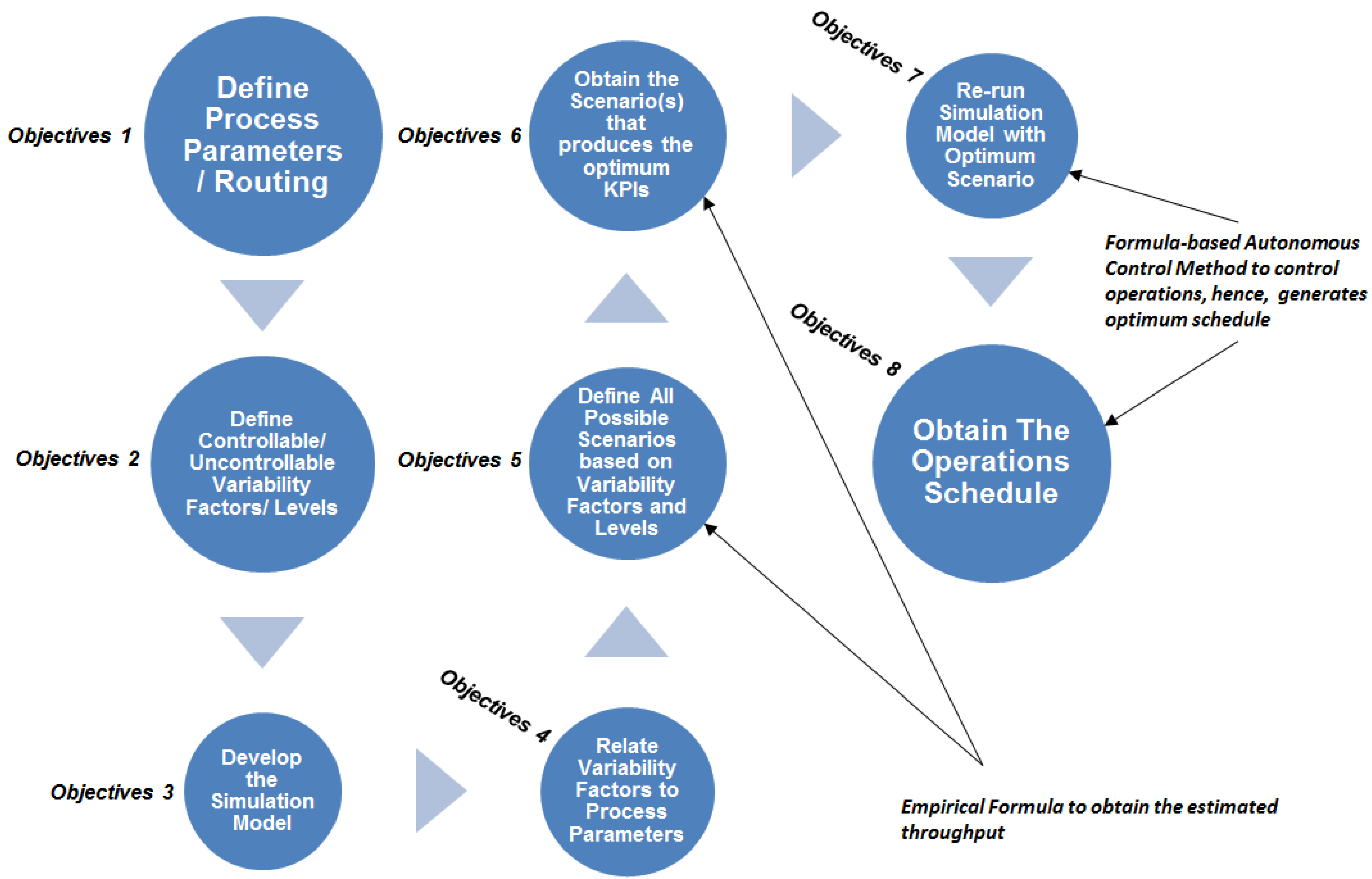


Figure 5.1: Research Relationship to the InnovateUK-funded project

The project runs on M1 motorway to transform it into a '*smart motorway*' and it involves construction of a central reservation barrier (CRB) project between junction 28 and 31 (Figure 5.2). Motorways in the UK are divided into junctions, and the junction is a set of links. The area between Junction 28 and 31 consists of four links. The concrete deliveries are the core driver for this construction project. The concrete supplier batches concrete to the construction site from two concrete plants at different locations. The concrete trucks drive to the construction site through the motorway. Once they reach the site, they undergo some site delays due to other site works. At the site, the concrete slump test is done and based on results, three possibilities can occur:

i. water is added to the load;
ii. truck is placed in a queue while waiting for the load to dry; or
iii. the truck proceeds to the discharge process if the extruder machine is free.

Once the load is discharged, the operation is considered complete. A saw cut process of the barrier takes place after the discharge process, however, it happens independently, so it does not affect the completed barrier length.

The real-world case study resembles a flexible flow line with large-volume steady production of medium-variety products, i.e., six sizes of concrete load batched from two plants. The flexibility here is generated mainly from human-dependent processes instead of flexible machines. The industrial partner of this project depends on the normal distribution to represent the process variability. Multiple variables within this project were identified that cause disruption to the concrete deliveries and accordingly the completed barrier length, i.e., throughput rate. This research is concerned with the variability within the part of the supply chain between concrete plants and the project site, where waste is usually anticipated, as a result of the lack of synchronisation of concrete deliveries. The waste in this construction site can be identified as either the number of trucks arriving on the construction site at same time, which can affect the concrete quality and causes unnecessary queues or no truck arrived, i.e., time lost while waiting for the concrete.

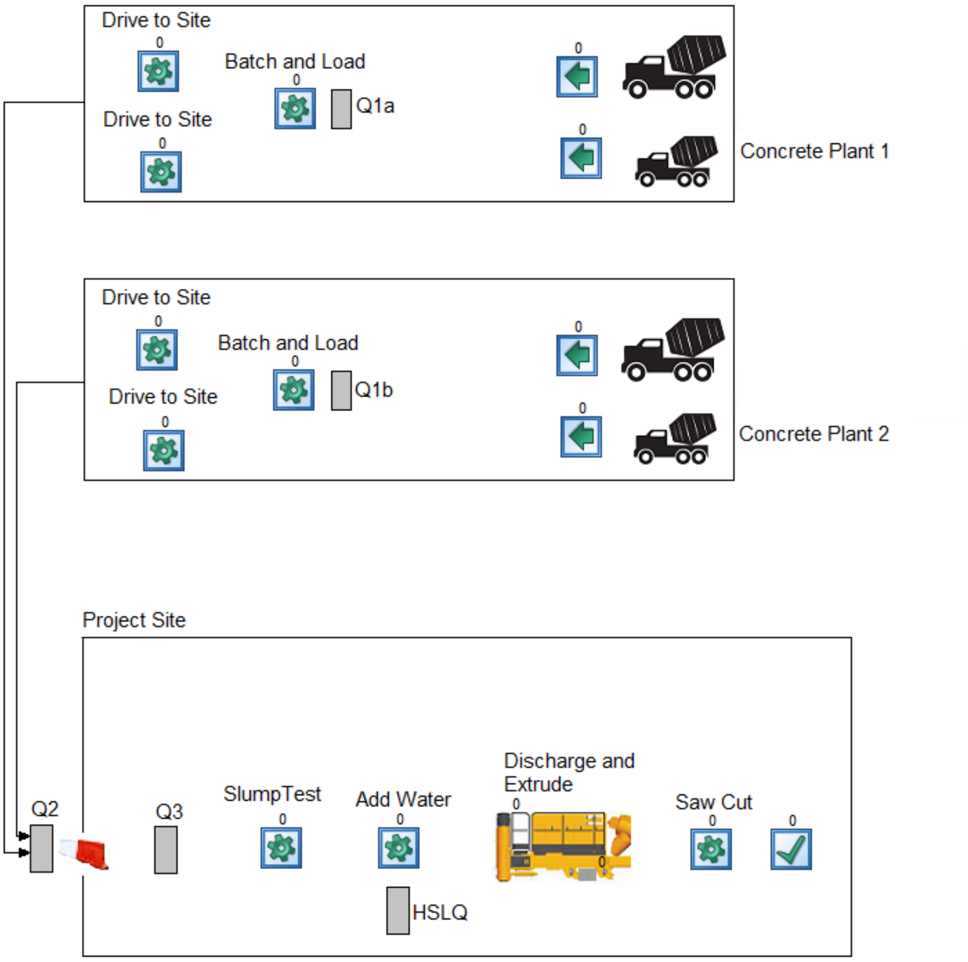


Figure 5.2: Model Representation of the Real-world Case Study

The research concern is to accurately estimate the anticipated throughput rate using the developed formula evaluative model, taking into consideration the variability factors that play a part in the construction operations and the different constraints and operational conditions during the working day, e.g., traffic congestion. The developed formula-based autonomous control method was then validated with this case study by using it to decide the arrival rates of the trucks and the operational setting for the other decision variables to synchronise the dispatch timing of trucks and minimise queues at the project site.

5.3 Research Methods and Steps

5.3.1 Data Collection

Collection of data related to the research from the industrial partner within the InnovateUK (Grant No. 18834-132285) research project. Data were collected through several on-site meetings and conference calls with the industrial partner and their sub-contractors. The data included the following:

i. detailed mapping of processes;
ii. operational constrains;
iii. resources, i.e., people, machines, equipment, vehicles, tools, etc.;
iv. current planning strategy; and
v. performance reports and records.

The performance report, i.e., concrete pour records, included the following information:

i. Date;
ii. Concrete Plant ID;
iii. Delivery Number;
iv. Concrete Load Size;
v. Batch Time;
vi. Time on Site;
vii. Start Discharge Time; and
viii. Finish Discharge Time.

5.3.2 Data Pre-processing

5.3.2.1 Data Sampling

The first four categories of collected data in Section 5.3.1 were used to define the underlying logic within the simulation model to replicate the actual operations at the construction project. The fifth category, i.e., Performance Reports, was used as the primary source of data to define the variability within the construction project. The processed performance reports are '*concrete pour records*' for M1 (Junction 28 to 31) CRB construction project by Costain, from 05th February 2014 to 02nd April 2014.

To minimise measurement or sampling bias, which can be caused by hidden variability parameters or insufficient data, the deterministic and stochastic factors and their levels of variability, as identified from these reports, were verified with the industrial partner.

As a result, another factor, the '*traffic congestion*' was identified. This factor has a direct effect on the delivery time of concrete truck which can be variable based on the time-of-the-day.

5.3.2.2 Data Transformation

The collected concrete pour records were transformed to define normally-distributed processing times that represent variability within the construction project:

i. Concrete batch time;
ii. Delivery time;
iii. Load conditioning time; and
iv. Discharge Time.

The mean and coefficient of variation of the time taken to batch the concrete to the truck was provided as $\mu_b = 2\,\text{min}/m^3$ and $c_b = 0.25$ by the industrial partner and verified during a meeting with the process owner, i.e., contractor. The delivery, load conditioning and discharge times were calculated from the concrete pour records as follows:

i. $Delivery\_Time = Time\_on\_Site - Batch\_Time$;
   $Load\_Conditioning\_Time = Start\_Discharge\_Time - Time\_on\_Site$;
   (Condition: Waiting time is not due to queuing); and
ii. $Discharge\_Time = Finish\_Discharge\_Time - Start\_Discharge\_Time$.

For the delivery time, the normal distribution, i.e., average and standard deviation of the truck delivery times μ_{del} and σ_{del}, from the two concrete plants, P1 and P2, were determined based on:

i. time-of-the-day;
ii. segmented times-of-the-day; and
iii. overall.

The data variations $v_{\mu_{del}}$ and $v_{\sigma_{del}}$ for each category were then examined to determine the suitable delivery time distribution. Based on the results shown in Table 5.1, the segmented times-of-the-day gave the minimal data variations, hence, it was chosen to define the truck delivery time distribution based on the time-of-the-day, as shown in Table 5.2.

In terms of load conditioning, three load conditions are used:

i. Suitable load: ready to be discharged when the extruder machine is available;

ii. High slump: the truck needs to wait for the concrete load with high water content to dry $(waiting\_time > 20\,\text{min})$; and

iii. Low slump: water is added to the load to adjust concrete properties $(waiting\_time \leq 20\,\text{min})$.

The instances of each load condition were extracted from the concrete pour records according to the waiting and queuing times, where a load can be clearly assigned to a specified category. Results were then plotted as shown in Figure 5.3.

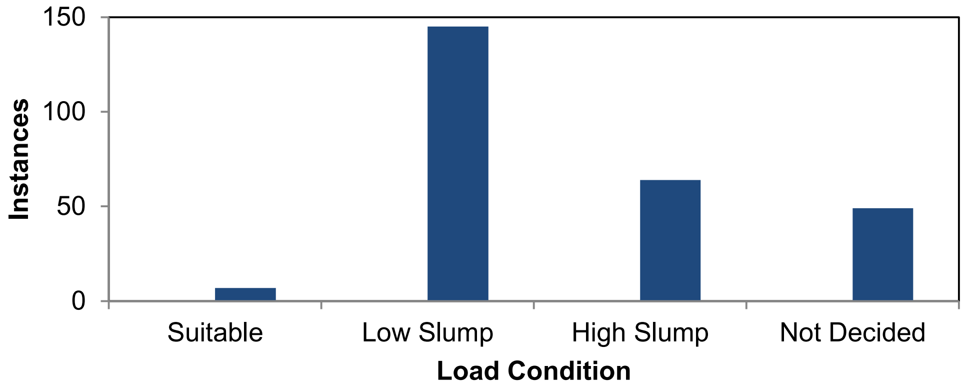


Figure 5.3: Histogram of Load Condition

Table 5.1: Data Variation for Three Categories of Delivery Time Distributions

	P1		P2	
	$v_{\mu_{del}}$	$v_{\sigma_{del}}$	$v_{\mu_{del}}$	$v_{\sigma_{del}}$
Time-of-the-day	7.29	6.99	15.39	22.35
Segmented times-of-the-day	**1.87**	**5.89**	**6.90**	**3.85**
Overall	16.37	14.09	33.57	12.37

Table 5.2: Delivery Time Distribution based on the Segmented Time-of-the-day

	P1		P2	
Time-of the-day	μ_{del} *(min)*	σ_{del} *(min)*	μ_{del} *(min)*	σ_{del} *(min)*
07:00	34.10	8.86	51.80	26.68
08:00	38.03	14.00	50.27	11.90
09:00-15:00	30.55	8.69	48.96	11.79
16:00	29.44	7.23	39.00	13.53

Normal distribution patterns of the load conditioning time μ_c and σ_c were then generated for the waiting time to condition the '*Low Slump*' and '*High Slump*' concrete loads (Table 5.3).

Table 5.3: Load Conditioning Time Distribution

Load Condition	μ_c *(min)*	σ_c *(min)*
Low Slump	11.14	5.45
High Slump	36.47	11.24

Finally, average and standard deviation of discharge time μ_{dis} and σ_{dis}, of each load size were determined, as shown in Table 5.4, and then the distribution per $m3$ of concrete load was calculated as ${}^{\mu_{dis}}\!/_{m^3} = 4.39\,\text{min}$ and ${}^{\sigma_{dis}}\!/_{m^3} = 1.53\,\text{min}$.

Table 5.4: Discharge Time Distribution for each Load Size

Load Size	μ_{dis} *(min)*	σ_{dis} *(min)*
5.5	23.00	8.47
6.0	24.85	9.60
7.5	32.57	11.04
8.0	39.30	12.10

5.3.2.3 Data Output – Simulated Throughput Rate

The output of the historical data for the real-world case study was obtained using simulation. Simulation models were developed based on the variability scenarios represented by the real-world case study using Simul8 simulation package (Figure 5.2).

The main elements of the model are:

i. Work Entry Point: where trucks enter the system before any processing operation is initiated;
ii. Batch and Load Queue (Q1): the queue of concrete trucks waiting to be batched;
iii. Batch and Load Process: the first process at the contractor concrete plant sites, where trucks are batched and loaded with concrete;

iv. Drive to Site Process: the second process, where trucks are delivering concrete to the construction site;
v. Site Access Delay Queue (Q2): delay to the concrete deliveries at the site access.
vi. Site Queue (Q3): trucks waiting to be load tested at the site.
vii. Slump Test Process: the slump test is applied to the concrete load to determine its suitability;
viii. Add Water Process: in case of dry load;
ix. High Slump Load Queue (HSLQ): where trucks wait for the high slump load to dry;
x. Discharge and Extrude Process: the only value-added process, where the load is being discharged at site;
xi. Saw Cut Process: a supplementary process that occur after the barrier is extruded; and
xii. Work Exit Point: where items leave the system.

Rules were defined in the simulation model to identify the variability and constraints of the operations.

The simulation model was programmed to loop through the scenarios and determine the steady state simulated throughput rate for each variability scenario using the '*Infinity*' method described in Section 4.5.2.2.

5.3.2.4 Data Cleaning

To reduce the noise generated by simulation modelling, Robinson (2004) confidence interval method was applied to the simulation models to determine the steady state simulated throughput rate with a confidence interval of 95%.

5.4 Summary

This chapter covered the specific research methods and steps of the real-world case study with the construction industry as part of the InnovateUK (Grant No. 18834-132285) research project. The collected data feed the validation sections of the Data Mining and Autonomous Control Frameworks as demonstrated with results in Chapter 6 and 7.

6 DATA MINING FRAMEWORK

6.1 Introduction

Chapter 4 presented the research methodology including the research design and an established Data Mining Framework for the different phases that were investigated during this study to relate the identified variables to the dependent variable of interest, i.e., throughput rate. It also presented the detailed methods and steps of data collection, pre-processing and analysis for the purpose of development of the formula evaluative model presenting the relationships between variables.

This chapter gives the results of these development phases and analyses the findings and implications that can be derived from the results.

6.2 Phase I – Data Pre-processing

Figure 6.1 shows the simulation output with a single run and multiple runs that satisfy the 95% confidence interval condition, according to Robinson (2004) method for a randomly selected scenario from the data set $I/II-A-6$. Apparently, the confidence interval stabilised the simulation output as shown in Figure 6.1a. However as shown in Table 6.1, the Mean Absolute Percentage Error (MAPE) does not necessary mean that there will always be an error in the throughput rate when running the simulation model with 0% CI since at some scenarios the error was zero, i.e., when Minimum Absolute Error (MINAE) equals zero. It implies only that that the multi runs will produce more robust and stable figures of the throughput rate which is backed by the visual results (Figure 6.1a). From the correlation analysis between the main parameters of the flow line and the error with a single run, i.e., 0% CI, in Table 6.2, it can be observed that the c_{av} plays a major role in this error.

The effect of warm-up (transient) state on the average throughput rate was then investigated. Figure 6.1b shows that for the selected scenario from the data set $I/II-A-6$, the increase in throughput rate over time becomes insignificant after 100,000 simulation time units especially with multiple iterations, i.e., less than 10^{-10}; hence, $t_{sim} \to \infty$; this period can be called the '*Saturation Period*'.

The results suggest that after a saturation period, though the throughput rate in the case of not defining a warm-up period is still marginally increasing (Figure 6.2a), the difference between throughput rate with and without a pre-defined warm-up period becomes negligible, i.e., $< 1.5 * 10^{-8}$ (Figure 6.2b). Based on time series graphical inspection and variance plots of several experiments over the range of the eight data sets ($I/II - A - 1$ and $I/II - A - 8$), the steady state response after the defined '*Saturation Period*' of any serial flow line that follows the rules of normal distribution can be approximated to:

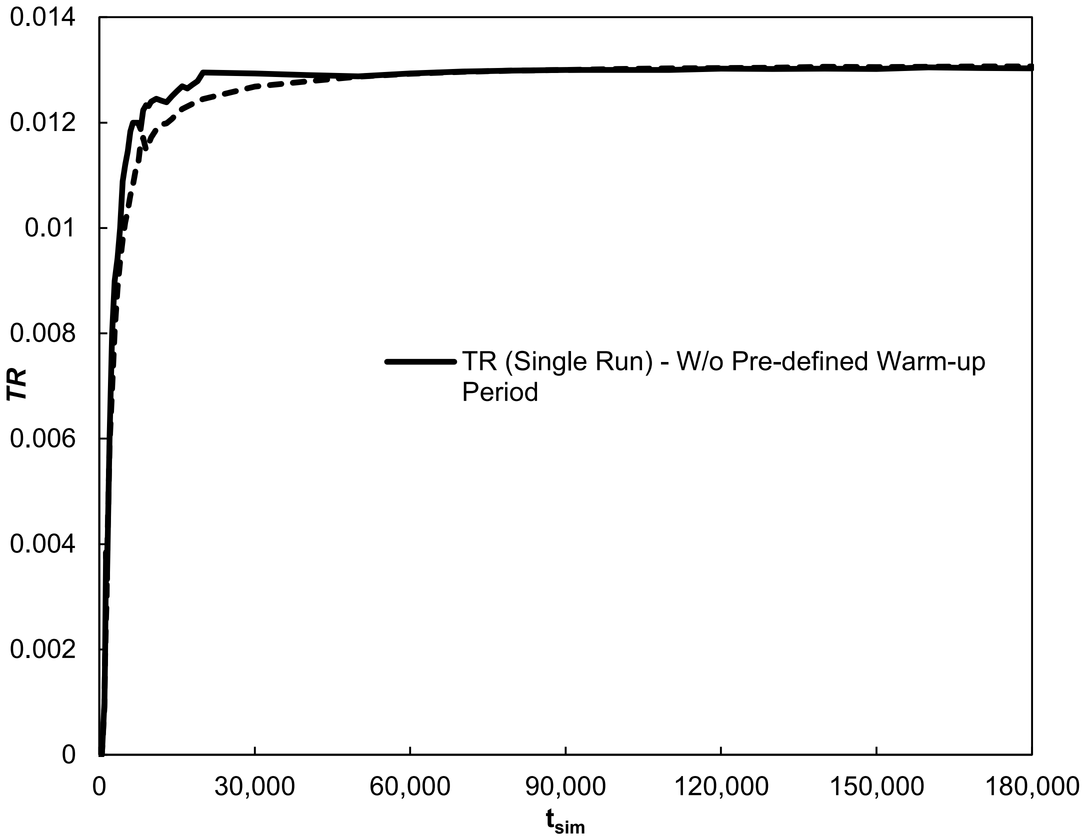
$$\lim_{t_{sim} \to \infty} f(t) = 55 N \mu_{\max} \tag{6.1}$$

Table 6.1: Error in the Throughput Rate due to the Confidence Interval

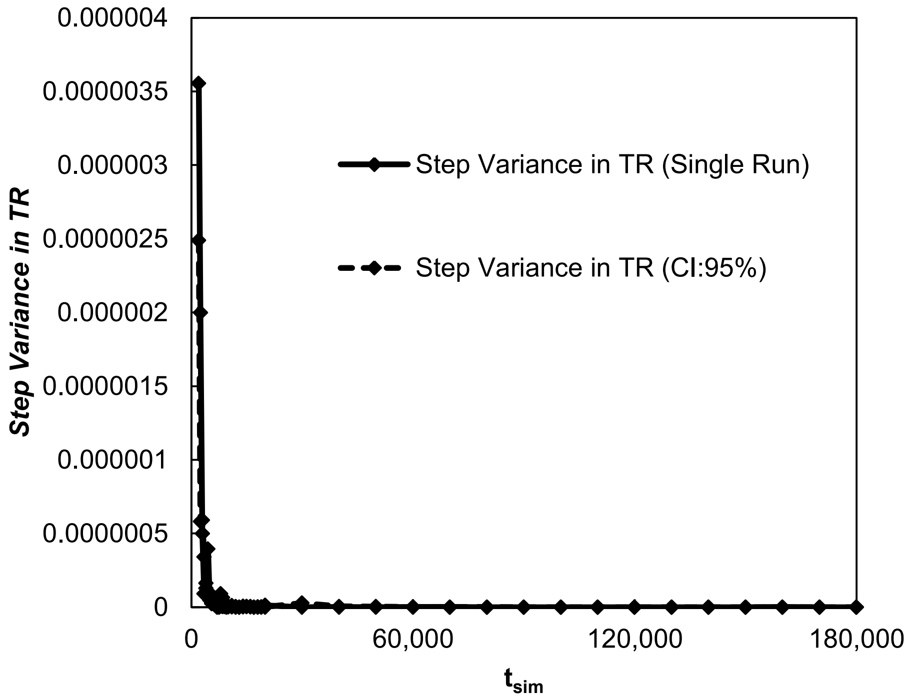
Statistical Measure	*Value*	*Statistical Measure*	*Value*
Mean Absolute Percentage Error (MAPE)	**0.6%**	Mean Absolute Error (MAE)	**0.0005**
Minimum Absolute Percentage Error (MINAPE)	**0%**	Minimum Absolute Error (MINAE)	**0**
Maximum Absolute Percentage Error (MAXAPE)	**7%**	Maximum Absolute Error (MAXAE)	**0.009**

Table 6.2: Correlation Analysis between the Flow Line-based Parameters the MAPE due to Confidence Interval

	$\mu_{\min}$	μ	$\mu_{\max}$	c	c_{av}	N
MAPE	-0.0961	-0.1514	-0.1047	0.0135	**0.6566**	-0.0425

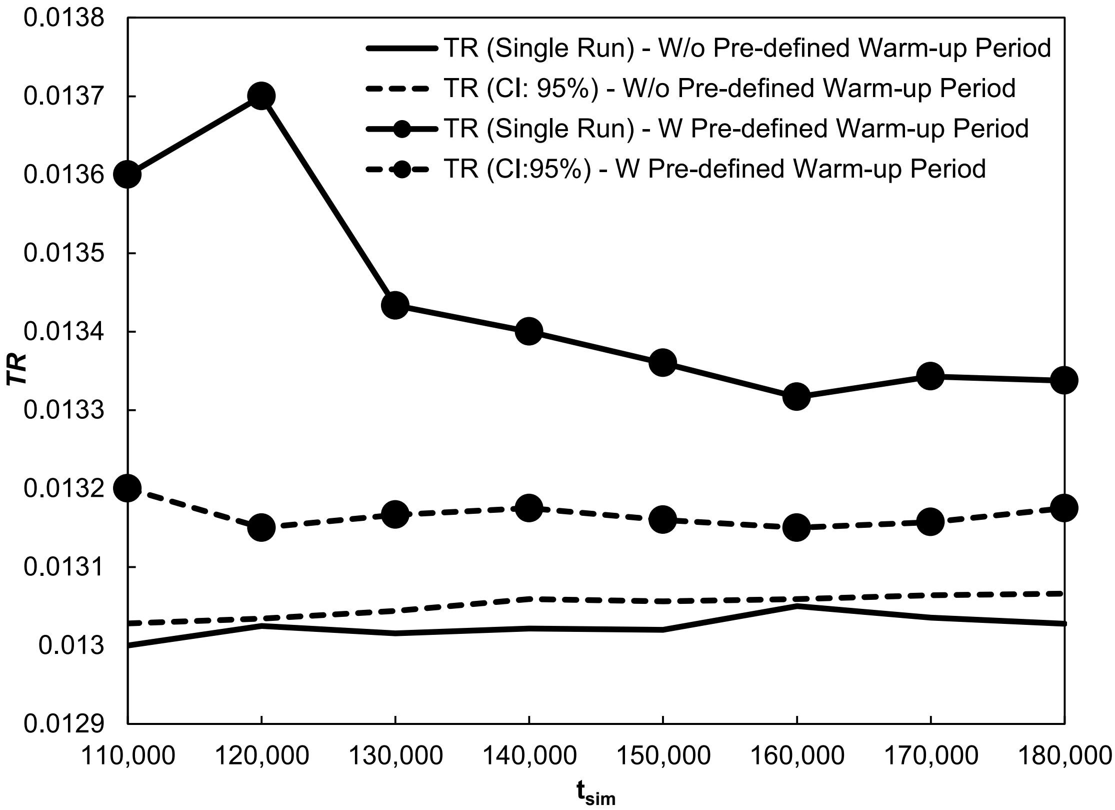


(a)

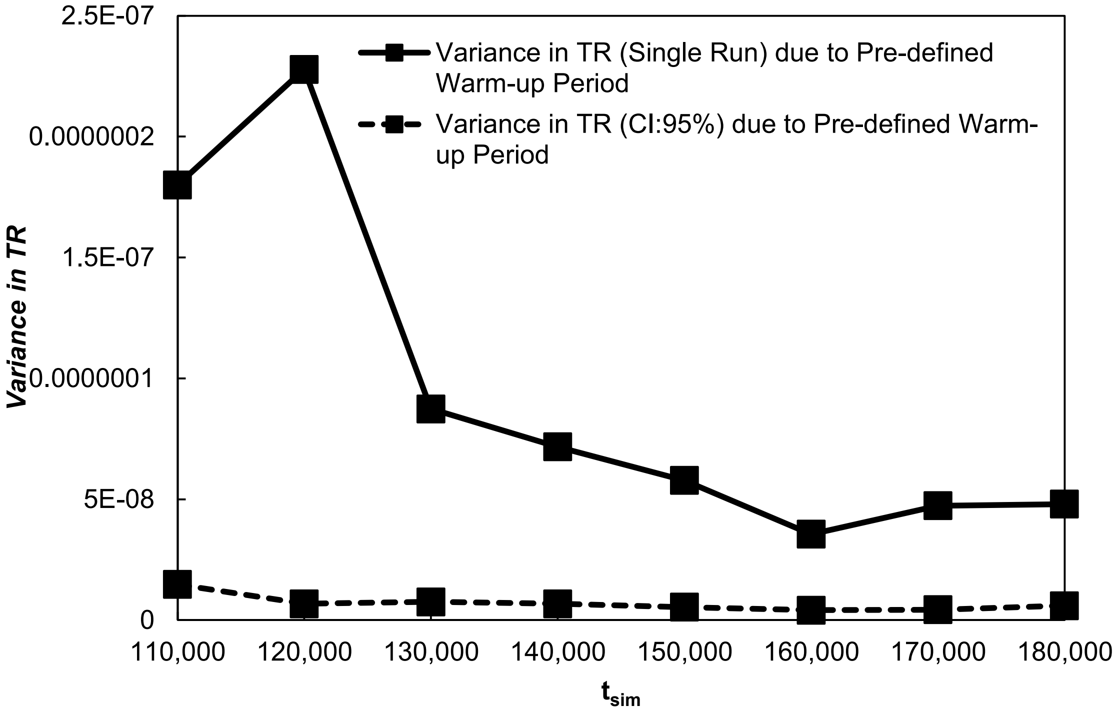


(b)

Figure 6.1: Time Series Graphical Inspection of (a) Throughput Rate and (b) Step Variance in Throughput Rate without Predefined Warm-up (Transient) Period



(a)



(b)

Figure 6.2: Time Series Graphical Inspection of (a) Throughput Rate and (b) Variance in Throughput Rate after Saturation Period with and without Pre-defined Warm-up (Transient) Period

6.3 Phase II – Feature Selection

The line-based parameter terms identified in the research design (Section 4.3) are examined here to select the predictors (features) to be used in the formula evaluative model building phase, Phase III, for asynchronous and the special case of synchronous flow lines. The results start with the general case, i.e., line based parameters for asynchronous flow lines, before converting the selected parameter terms to process-based for the synchronous flow line case.

6.3.1 Average Coefficient of Variation (c_{av})

As mentioned in Section 2.6.2, c_{av} is the root cause of disturbance in TR. Therefore, the data set used for this parameter was constructed of three sub-sets used to test and validate the relationship with the throughput rates, $\mathbf{TR}_1, \mathbf{TR}_2, \mathbf{TR}_3$ for the three sub-sets, respectively, as follows:

i. $IV - c_{av} - 1$:

$$\mu_i = \begin{cases} 1 & for\, N \in \{1,2,\ldots,7\}, \\ 10 & for\, N = 8, \\ 1 & for\, N \in \{9,10,\ldots,15\}, \end{cases} \qquad c_i \in \{0,0.01,0.025,0.05,0.075,0.1,0.25,0.5,0.75,1\}, N = 15, l = 7$$

ii. $IV - c_{av} - 2$:

$$\mu_i = \begin{cases} 5 & for\, N \in \{1,2,\ldots,7\}, \\ 10 & for\, N = 8, \\ 5 & for\, N \in \{9,10,\ldots,15\}, \end{cases} \qquad c_i \in \{0,0.01,0.025,0.05,0.075,0.1,0.25,0.5,0.75,1\}, N = 15, l = 7$$

iii. $IV - c_{av} - 3$:

$$\mu_i \in \{1,2,\ldots,60\}, c_i \in \{0,0.01,0.025,0.05,0.075,0.1,0.25,0.5,0.75,1\}, N = 15, l = 7$$

TR exhibited very high negative correlation with average coefficient of variation c_{av} and the exponential term of coefficient of variation $e^{c_{av}}$ for the three sub-sets (Table 6.3 and Figure 6.3). Figure 6.7a shows the effect of the change in the throughput rate due to the change of the individual variability factor, i.e., c_{av}, only. The throughput rate was normalised to the throughput rate of the initial value of the variability factor TR^{in} for illustration purposes:

$$\mathbf{TR}^{norm} = \frac{\mathbf{TR}_{ss,h}}{\mathbf{TR}^{in}_{ss,h}} \tag{6.2}$$

where in stands for initial, ss for the sub-set and h is the sub-set number.

Though the exponential term has the highest correlation with TR, c_{av} shares almost the same significance with it which suggests that the throughput formula might include both terms of c_{av}. There was close consistency along the sub-sets, however, the relationship becomes stronger with the increase in randomness and mean processing time.

The significance analysis showed that the two terms of coefficient of variation are strong enough to represent the variability of the dependent response perfectly (*p*-value=0). The exponential term of coefficient of variation is highly correlated to the change in TR when all the other parameters are constant.

Finally, to determine if these terms can solely represent TR, the correlation coefficient of these two terms and TR was calculated for the whole data set IV, including sub-sets for all parameters. The results showed very weak relationships suggesting that these terms cannot solely represent TR.

Table 6.3: (i) Correlation and (ii) ANOVA Analysis of the Relationship between Coefficient of Variation Terms and the Throughput Rate

(i)

Term	TR_1	TR_2	TR_3
c_{av}	**-0.90**	**-0.97**	**-0.99**
c_{av}^{-1}	0.33	0.41	0.45
$\log c_{av}$	-0.66	-0.78	-0.83
$\log {}^{1}\!/_{c_{av}}$	0.66	0.78	0.83
$e^{c_{av}}$	**-0.95**	**-0.99**	**-0.99**
$e^{c_{av}^{-1}}$	0.17	0.22	0.25

(ii)

Source	*DF*	*Adj SS*	*Adj MS*	*f-value*	*p-value*
Regression	2	0.000001	0	1256.59	0
c_{av}	**1**	**0**	**0**	**253.38**	**0**
$e^{c_{av}}$	**1**	**0**	**0**	**468.27**	**0**
Error	5	0	0		
Total	7	0.000001			

6.3.2 Location Ratio of the Process with Maximum Mean Processing Time (l)

The data set used for the location ratio of the process with maximum mean processing time was chosen to examine each individual location of the flow line as follows:

$$IV - l : \mu_i = \begin{cases} 1, & for\, i \neq l, \\ 10, & for\, i = l, \end{cases} \quad c_i = 0.1, N = 15, l \in \{1,2,...,15\}$$

As shown in Table 6.4 and Figures 6.4 and 6.7b, the location and number of the processes with maximum mean processing time has no correlation with TR.

Since there is no dependency between the location of the process with maximum mean processing time and TR, this parameter is of no use to the prediction model.

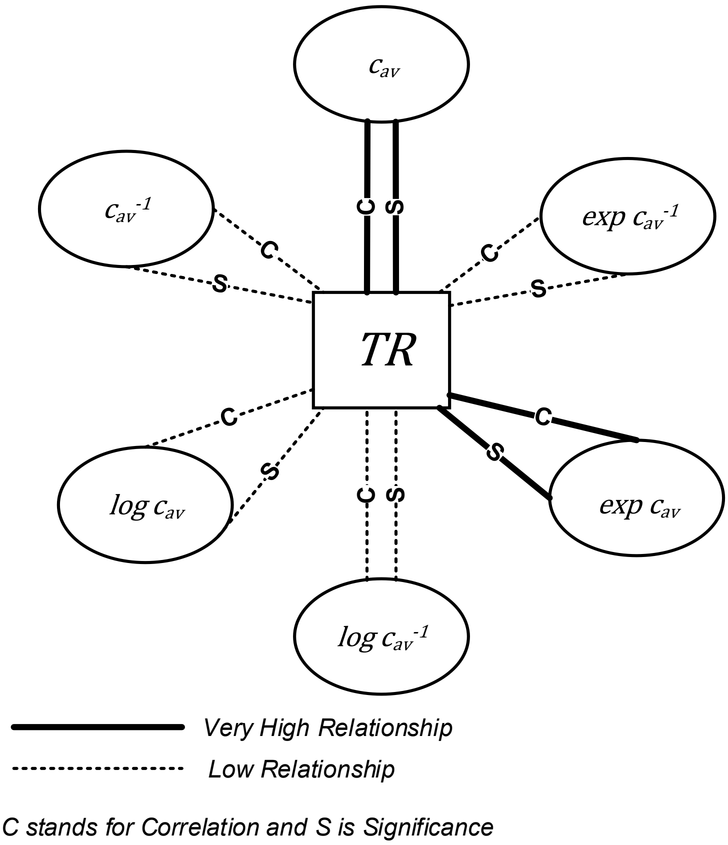
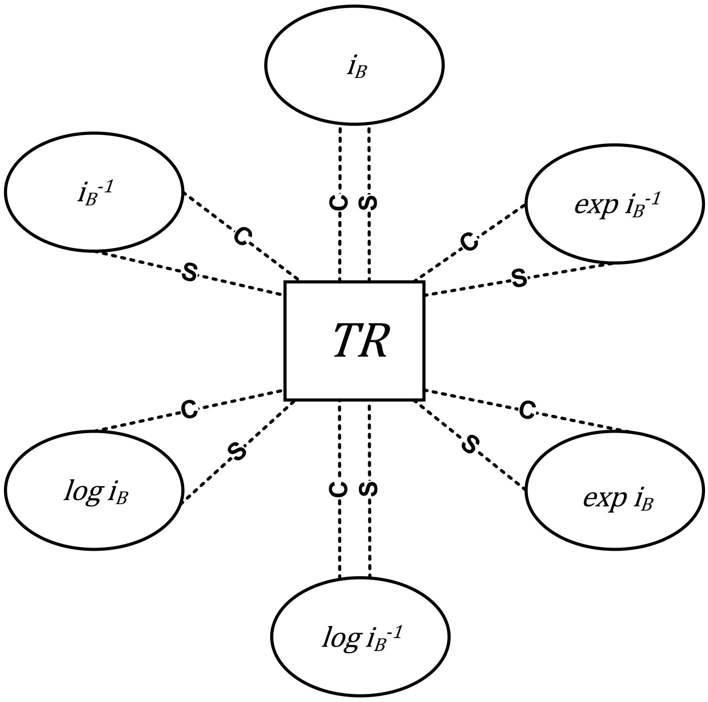


Figure 6.3: Relationship Diagram for Coefficient of Variation Terms

Table 6.4: Correlation Analysis of the Relationship between Location Ratio of the Process with Maximum Mean Processing Time Terms and the Throughput Rate

Term	TR_1
l	-0.06
l^{-1}	-0.12
$\log l$	0.00
$\log \frac{1}{l}$	-0.00
e^{l}	-0.07
$e^{l^{-1}}$	-0.21



------------------ *Low Relationship*

C stands for Correlation and S is Significance

Figure 6.4: Relationship Diagram for Location Ratio of the Process with Maximum Mean Processing Time Terms

6.3.3 Maximum Mean Processing Time within Flow Line $(\mu_{\max})$

As for the maximum mean processing time which is the sole representative of variability in the throughput rate when $c_{av} = 0$, it is not always the only defining line-based parameter. When c_{av} increases, several parameters start to affect TR.

The data set for this parameter included the following two sub-sets for flow lines with different lengths:

$$IV - \mu_{\max} - 1 : \mu_i = \begin{cases} 1, & for\, i \in \{1,2\}, \\ \mu_{\max}, & for\, i = 3, \end{cases} \quad \mu_{\max} \in \{2,3,\ldots,10\}, c_i = 0.1, N = 3, l = 3$$

$IV - \mu_{\max} - 2$:

$$\mu_i = \begin{cases} 1, & for\, i \in \{1,2,\ldots,8\}, \\ \mu_{\max}, & for\, i \in \{9,10,\ldots,15\}, \end{cases} \quad \mu_{\max} \in \{2,3,\ldots,10\}, c_i = 0.1, N = 15, l \in \{9,10,\ldots,15\}$$

All terms apart from exponential term showed a high correlation to *TR* (Table 6.5). However the multiplicative inverse $\left(\mu_{\max}^{-1}\right)$ is still the predominant in terms of correlation and significance as shown in Figure 6.5 (*f*-value=4640 and *p*-value=0).

$\mu_{\max}^{-1}$ is strong enough to represent the variability of the dependent response TR perfectly for the two sub-sets. Subsequently, this term also exhibits very high correlation across Class IV data set for all parameters (Figure 6.7c).

Table 6.5: (i) Correlation and (ii) ANOVA Analysis of the Relationship between Maximum Mean Processing Time Terms and the Throughput Rate

(i)			(ii)					
Terms	TR_1	TR_2	*Source*	*DF*	*Adj SS*	*Adj MS*	*f-value*	*p-value*
$\mu_{\max}$	**-0.90**	**-0.90**	*Regression*	1	0.1364	0.1364	21533270	0
$\mu_{\max}^{-1}$	**1.00**	**1.00**	$\mu_{\max}^{-1}$	**1**	**0.1364**	**0.1364**	**21533270**	**0**
$\log \mu_{\max}$	**-0.97**	**-0.97**	*Error*	7	0	0		
$\log \frac{1}{\mu_{\max}}$	**0.97**	**0.97**	*Total*	8	0.1364			
$e^{\mu_{\max}}$	-0.49	-0.49						
$e^{\mu_{\max}^{-1}}$	**1.00**	**1.00**						

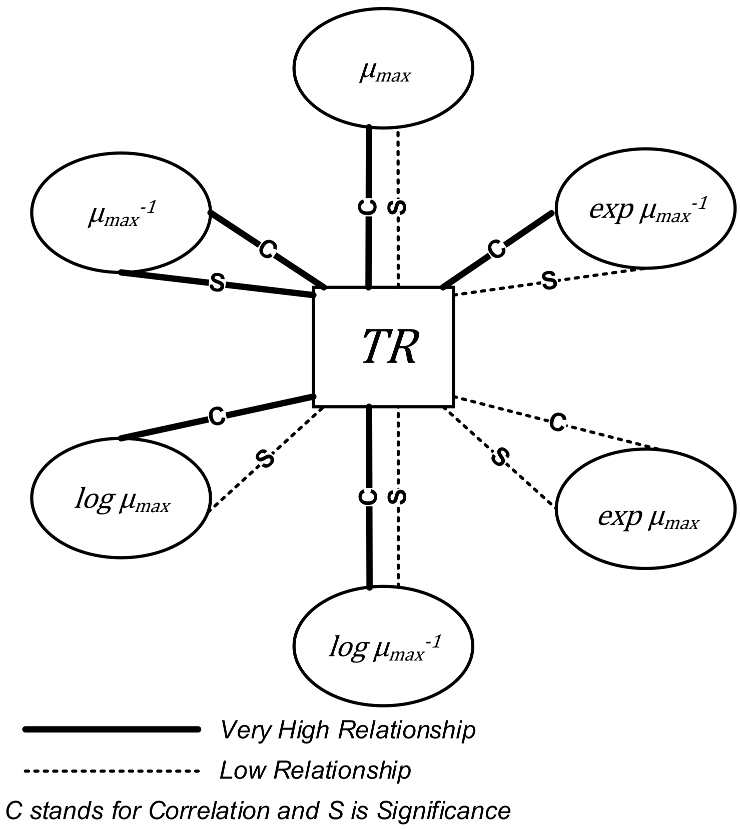


Figure 6.5: Relationship Diagram for Maximum Mean Processing Time Terms

6.3.4 *Length* (N)

For the length, the sub-sets were designed, as mentioned in 4.5.3.1, to keep other parameters constant and μ and c as stable as possible. Nine sub-sets were used with variable $\mu_{\max}$ of the flow line as follows:

$$IV-N-1i:\mu_i=\begin{cases}1, & for\ i\le 0.5N,\\ \mu_{\max}, & for\ i>0.5N,\end{cases}\quad \mu_{\max}=2, c_i=0.1, N\in\{4,6,8,10,12\}$$

$IV-N-1ii:$

$$\mu_i=\begin{cases}1, & for\ i\le 0.538N,\\ \mu_{\max}, & for\ i>0.538N,\end{cases}\quad \mu_{\max}=2, c_i=0.1, N\in\{13,15,17,19,21,23,25,27,29\}$$

$$IV-N-2i:\mu_i=\begin{cases}1, & for\ i\le 0.5N,\\ \mu_{\max}, & for\ i>0.5N,\end{cases}\quad \mu_{\max}=3, c_i=0.1, N\in\{4,6,8,10,12\}$$

$IV-N-2ii:$

$$\mu_i=\begin{cases}1, & for\ i\le 0.538N,\\ \mu_{\max}, & for\ i>0.538N,\end{cases}\quad \mu_{\max}=3, c_i=0.1, N\in\{13,15,17,19,21,23,25,27,29\}$$

…

$$IV-N-10i:\mu_i=\begin{cases}1, & for\ i\le 0.5N,\\ \mu_{\max}, & for\ i>0.5N,\end{cases}\quad \mu_{\max}=10, c_i=0.1, N\in\{4,6,8,10,12\}$$

$IV-N-10ii:$

$$\mu_i=\begin{cases}1, & for\ i\le 0.538N,\\ \mu_{\max}, & for\ i>0.538N,\end{cases}\quad \mu_{\max}=10, c_i=0.1, N\in\{13,15,17,19,21,23,25,27,29\}$$

Results for the nine sub-sets are in close agreement, hence, only the first four sub-sets are shown in Table 6.6 and Figures 6.6 and 6.7d. The length terms, except for the exponential term, are highly correlated to TR. N^{-1} exhibits the highest correlation; however N explains some of the variability in TR and the two terms together are strong enough to fully represent, at different degrees of significance, the variability in TR due to the change in length (*f*-value=4925.5 and *p*-value=0).

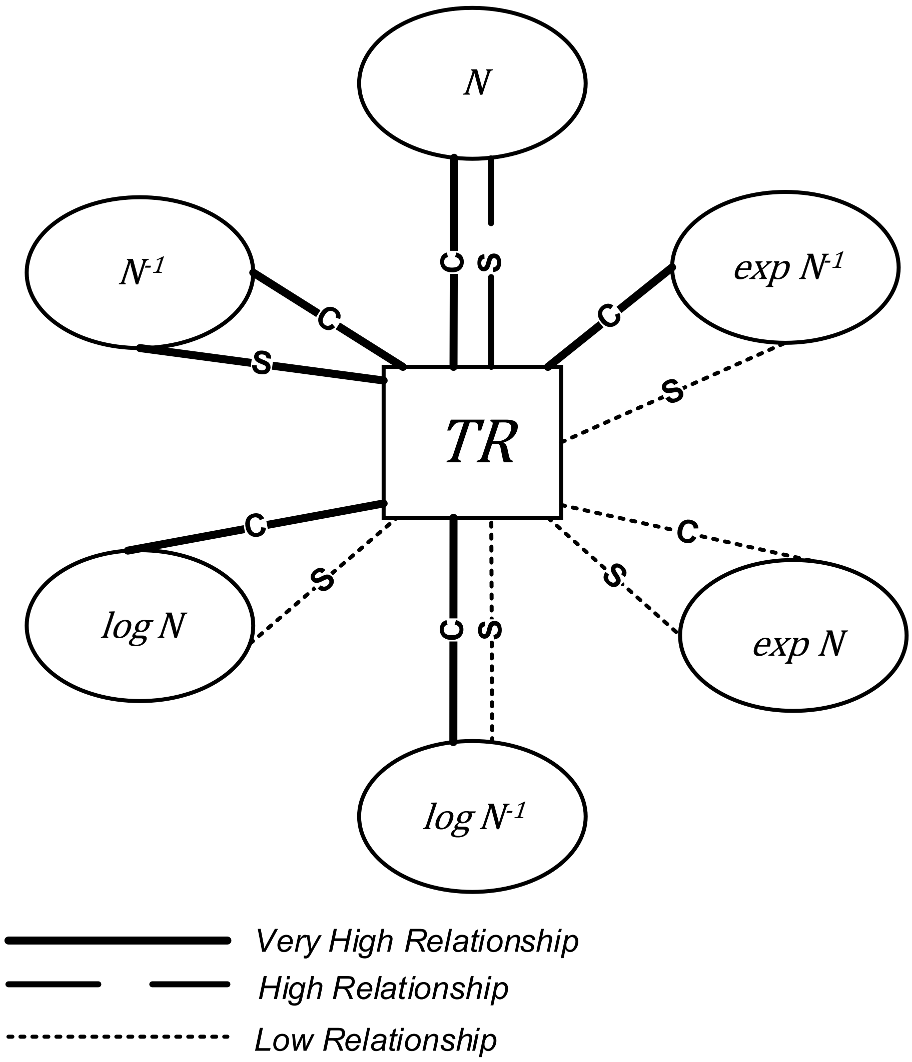
Table 6.6: (i) Correlation and (ii) ANOVA Analysis of the Relationship between Length Terms and the Throughput Rate

(i)

Terms	TR_1	TR_2	TR_3	TR_4
N	**-0.87**	**-0.88**	**-0.88**	**-0.89**
N^{-1}	**1.00**	**1.00**	**1.00**	**1.00**
$\log N$	**-0.97**	**-0.97**	**-0.97**	**-0.98**
$\log \frac{1}{N}$	**0.97**	**0.97**	**0.97**	**0.98**
e^{N}	-0.30	-0.30	-0.32	-0.33
$e^{N^{-1}}$	**1.00**	**0.99**	**0.99**	**0.99**

(ii)

Source	*DF*	*Adj SS*	*Adj MS*	*f-value*	*p-value*
Regression	3	0.0034	0.0011	4925.5	0
N	**1**	**0.000002**	**0.000002**	**8.24**	**0.017**
N^{-1}	**1**	**0.00004**	**0.00004**	**149.8**	**0**
$\log \frac{1}{N}$	1	0.000001	0.000001	3.59	0.087
Error	10	0.000002	0		
Total	13	0.003443			



C stands for Correlation and S is Significance

Figure 6.6: Relationship Diagram for Length Terms

6.3.5 *Minimum, Average and Coefficient of Variation of Mean Processing Time within Flow Line* $(\mu_{\min}, \mu, c)$

Finally, for the minimum, average and coefficient of variation of mean processing time, and due to the inter-dependency of these parameters, two data sets were generated. The first data set has a constant $\mu_{\min}$ while the other has a steady $\mu_{\max}$. The two data sets were then divided into sub-sets as follows:

i. μ *and* c

$$IV-\mu,c-1:\mu_i=\begin{cases}1 & for\ N\in\{1,2,...,7\},\\ \mu_{\max} & for\ N=8,\\ 1 & for\ N\in\{9,10,...,15\},\end{cases}\qquad \mu_{\max}\in\{2,3,...,60\},c_i=0.75,N=15$$

$$IV-\mu,c-2:\mu_i=\begin{cases}\mu_{\max}, & for\ i<0.5N,\\ 1, & for\ i>0.5N,\end{cases}\quad \mu_{\max}\in\{2,3,...,60\},c_i=0.75,N=15$$

$$IV-\mu,c-3:\mu_i=\begin{cases}1, & for\ i\le 0.538N,\\ \mu_{\max}, & for\ i>0.538N,\end{cases}\quad \mu_{\max}\in\{2,3,...,60\},c_i=0.75,N=15$$

$$IV-\mu,c-4:\mu_i=\begin{cases}\mu_{\max} & for\ N\in\{1,2,...,7\},\\ 1 & for\ N=8,\\ \mu_{\max} & for\ N\in\{9,10,...,15\},\end{cases}\qquad \mu_{\max}\in\{2,3,...,60\},c_i=0.75,N=15$$

i. $\mu_{\min}$, μ *and* c

$$IV-\mu_{\min},\mu,c-1:\mu_i=\begin{cases}\mu_{\min} & for\ N\in\{1,2,...,7\},\\ 60 & for\ N=8,\\ \mu_{\min} & for\ N\in\{9,15\},\end{cases}\qquad \mu_{\min}\in\{2,3,...,59\},c_i=0.75,N=15$$

$$IV-\mu_{\min},\mu,c-2:\mu_i=\begin{cases}60, & for\ i<0.5N,\\ \mu_{\min}, & for\ i>0.5N,\end{cases}\quad \mu_{\min}\in\{2,3,...,59\},c_i=0.75,N=15$$

$$IV-\mu_{\min},\mu,c-3:\mu_i=\begin{cases}\mu_{\min}, & for\ i\le 0.538N,\\ 60, & for\ i>0.538N,\end{cases}\quad \mu_{\min}\in\{2,3,...,59\},c_i=0.75,N=15$$

$$IV-\mu_{\min},\mu,c-4:\mu_i=\begin{cases}60 & for\ N\in\{1,2,...,7\},\\ \mu_{\min} & for\ N=8,\\ 60 & for\ N\in\{9,10,....,15\},\end{cases}\qquad \mu_{\min}\in\{2,3,...,59\},c_i=0.75,N=15$$

The correlation and significance of these terms is ambiguous and complex. Results for these parameters change from one experiment to another as shown in Appendix E (P. A-11) for the subsets $IV-\mu_{\min},\mu,c-1$ to $IV-\mu_{\min},\mu,c-4$.

Hence, it is more difficult to associate the change in *TR* with specific terms of these three parameters. Therefore, to deal with the discrepancy and avoid neglecting any important relationships, only the terms with weak strength and did not show significance for a single data set were removed.

Each parameter term was given a score based on its relationship with *TR* with more emphasis given to significance over strength. Table 6.7 shows the correlation coefficient for each sub-set. As shown, the correlation is strong, i.e., higher than 0.8 (Section 4.3), for two sub-sets out of 8. Hence, the score given to this parameter term was 2/8. The same criterion was applied to the significance of the parameter term but the score was doubled. The total weighted score was then calculated and the pass score was set low, i.e., 25% or 6/24 (Table 6.8).

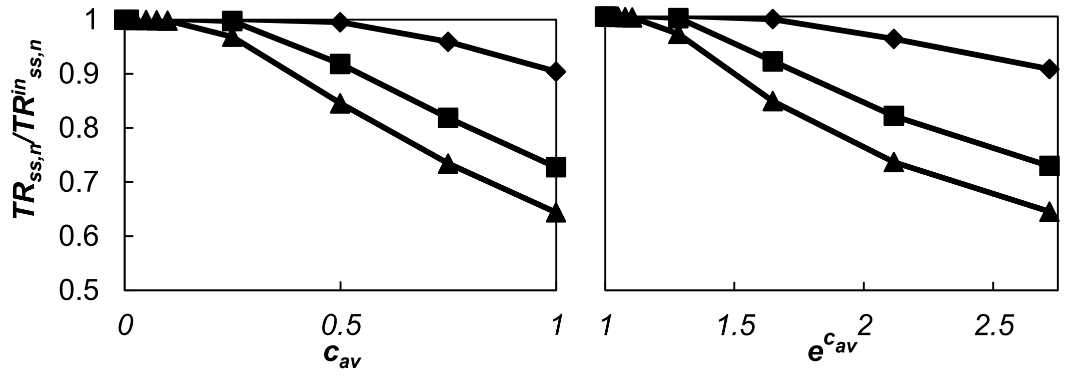
Hence, the removed terms were $e^{\mu}, e^{c}, e^{c^{-1}}, \mu_{\min}, \mu_{\min}^{-1}, \log \mu_{\min}, \log \frac{1}{\mu_{\min}}$ and $e^{\mu_{\min}}$.

Table 6.7: Relationship Score and Inclusion or Exclusion Decision for Minimum, Average and Coefficient of Variation of Mean Processing Time Terms

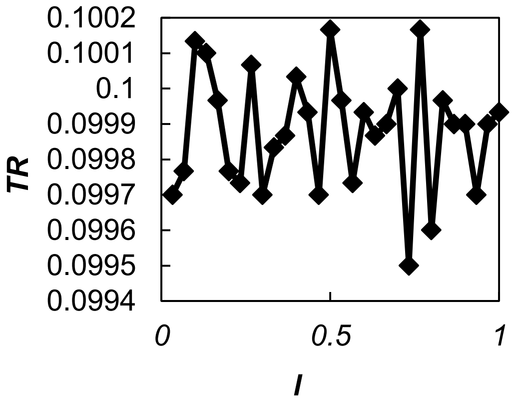
Predictor Terms	*Score - Strength (out of 8)*	*Score - Significance (out of 16)*	*Total Score (out of 24)*	*Decision*
μ	**2**	**4**	**6**	***Include***
μ^{-1}	**5**	**2**	**8**	***Include***
$\log \mu$	**6**	**2**	**8**	***Include***
$\log \frac{1}{\mu}$	**6**	**0**	**6**	***Include***
e^{μ}	3	0	3	*Exclude*
$e^{\mu^{-1}}$	**5**	**6**	**11**	***Include***
$\mu_{\min}$	4	0	4	*Exclude*
$\mu_{\min}^{-1}$	0	0	0	*Exclude*
$\log \mu_{\min}$	2	0	2	*Exclude*
$\log \frac{1}{\mu_{\min}}$	2	0	2	*Exclude*
$e^{\mu_{\min}}$	0	0	0	*Exclude*
$e^{\mu_{\min}^{-1}}$	**0**	**8**	**8**	***Include***
c	**5**	**4**	**9**	***Include***
c^{-1}	**6**	**8**	**14**	***Include***
$\log c$	**8**	**4**	**12**	***Include***
$\log \frac{1}{c}$	**8**	**2**	**10**	***Include***
e^{c}	1	0	1	*Exclude*
$e^{c^{-1}}$	0	0	0	*Exclude*

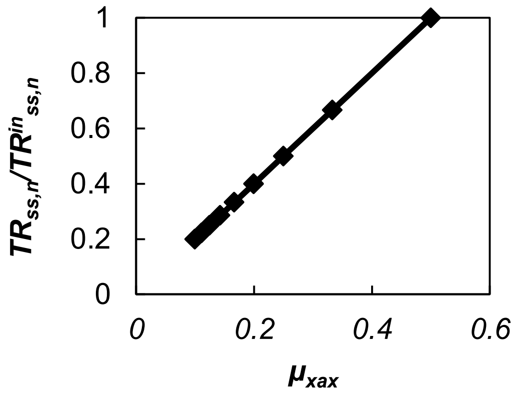
Table 6.8: Correlation Coefficient between the Direct Term of Average Mean Processing Time and the Throughput Rate of each Sub-set

	$TR_{ss,1}$	$TR_{ss,2}$	$TR_{ss,3}$	$TR_{ss,4}$	$TR_{ss,5}$	$TR_{ss,6}$	$TR_{ss,7}$	$TR_{ss,8}$
μ	**-0.99**	-0.75	-0.76	**-0.93**	-0.70	-0.67	-0.67	-0.67

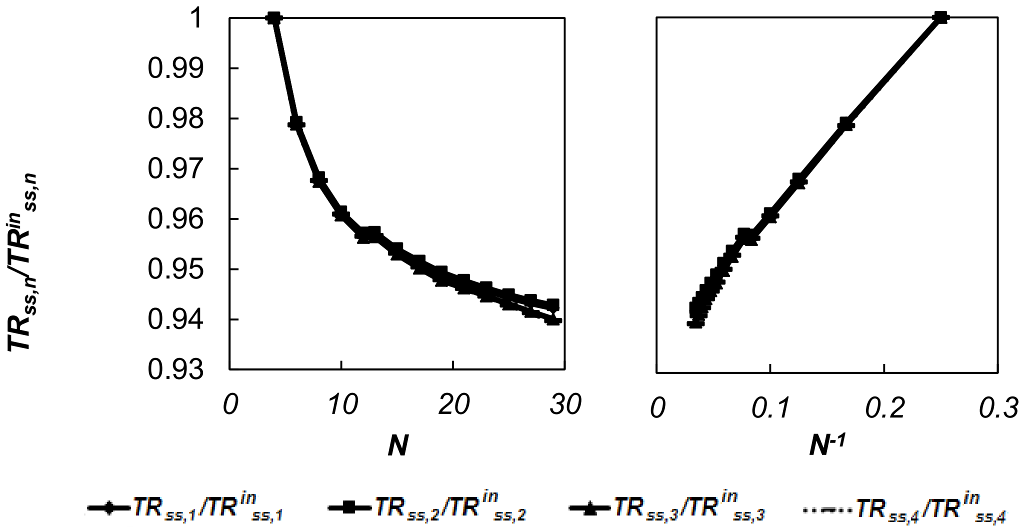


(a) Coefficient of Variation





(b) Location Ratio (c) Maximum Mean Processing Time



(d) Length

Figure 6.7: Change of Throughput Rate with the Most Important Terms of each Line-based Parameter

6.3.6 *Best Sub-set Regression*

To determine if the discrepant parameter terms are important to be included in the modelling step, the parameter terms will need to be examined if they add value to the model. Best sub-sets regression is a method that can initially be used to do this validation in one step since it will provide statistical measures for the best single-variable model, 2-variables and so on.

Best sub-set regression was applied to the full original data sets ($I/II-A-1$ to $I/II-A-8$) described in Section 4.5.2.1.2 with two conditions; one with the certain parameter terms that have an agreement throughout the different sub-sets and the second with the uncertain parameter terms that are inconsistent.

Results were analysed using the following statistical measures:

i. R^2;
ii. Adjusted R^2;
iii. Predicted R^2;
iv. Mallows's Cp; and
v. Standard error.

For the first condition that includes the following terms ${\mu_{\max}}^{-1}$, c_{av}, $e^{c_{av}}$, N and N^{-1}, results show that the five terms are important and they can give an accurate model with R^2 of 96.7% even without interactions and higher polynomial orders.

The results shown in Table 6.9a suggested that all these terms are needed to define the throughput rate using linear terms only; exclusion of any one has a significant effect on Mallows's Cp, which describes the prediction accuracy of the model with the selected predictors. Ideally, Mallows's Cp has to be equal the number of predictors plus one (for the constant); this condition was met when all predictors are included.

However, variability is still not fully defined (R^2 of 96.7%). Improvements can be approached statistically by investigating if;

i. there are more parameter terms that can be added; or
ii. determine the right term and order of the existing predictors.

In the second regression, the predictor terms from first condition were used as main predictors (included in all models) and the remaining following predictor terms were set as free predictors. Adding these predictors with complex relationship with *TR* (i.e., $\mu_{\min}, \mu$ and c) has improved the accuracy and significance of the model (Table 6.9b), however, the terms that did not show effect on the model and can be excluded from the forthcoming stages are $\log \frac{1}{\mu}$ and $\log \frac{1}{c}$.

Hence, free predictors, for use as supporting predictors during the model building stage, were reduced to $e^{\mu_{\min}^{-1}}$, μ, μ^{-1}, $\log\mu$, $e^{\mu^{-1}}$, c, c^{-1} and $\log c$.

These terms were converted from line-based to process-based for the synchronous flow line case by including the matching parameters, c, N and the dominant mean processing time which in this case is $\mu_{\max}$. Hence, the selected main predictors for the asynchronous flow line model translated in process-based parameter terms are c, e^{c}, μ^{-1}, N and N^{-1} with no supporting (free) predictors.

The results with the asynchronous case showed improvements in the explained relationship through statistical measures which reach the maximum when all the free predictors are included. However, the best sub-set regression model did not explain the relationship perfectly since the degrees of freedom in modelling using normal best sub-set regression technique are limited, i.e., linear, which was investigated during the model building stage.

Table 6.9: Best Sub-set Regression Analysis for (a) Main Predictors only and (b) Main and Free Predictors

(a)

Vars	R^2	R^2*(adj)*	*Mallows Cp*	*S*	${\mu_{\max}}^{-1}$	c	$e^{c_{av}}$	N	N^{-1}
1	89.2	89.2	262821	0.0149	X				
2	96.7	96.7	1646.1	0.0083	X		X		
3	96.7	96.7	748.2	0.0082	X	X	X		
4	96.7	96.7	100	0.008	X	X	X		X
5	96.7	96.7	81.1	0.008	X	X	X		X
6	**96.7**	**96.7**	**6**	**0.0082**	**X**	**X**	**X**	**X**	**X**

(b)

Vars	R^2	R^2*(adj)*	*Mallows Cp*	*S*	$e^{{\mu_{\min}}^{-1}}$	μ	μ^{-1}	$\log\mu$	$e^{\mu^{-1}}$	c	c^{-1}	$\log c$
1	97.3	97.3	2991	0.007			X					
2	97.3	97.3	2064	0.007			X			X		
3	97.3	97.3	908	0.007			X			X		X
4	97.3	97.3	727	0.007			X		X	X		X
5	97.3	97.3	200	0.007			X	X	X	X		X
6	97.3	97.3	149	0.007			X	X	X	X	X	X
7	97.3	97.3	103	0.007	X		X	X	X	X	X	X
8	97.3	97.3	58	0.007	X	X	X	X	X	X		X
9	**97.3**	**97.3**	**14**	**0.007**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**

X indicates the predictor is included in the regression model

6.4 Phase III – Model Building

Closed-form formulas were developed for the throughput rate of synchronous and asynchronous serial flow lines that follow the rules of normal distribution. The formulas are based on an empirical study using an integration of DES modelling and supervised machine learning techniques.

6.4.1 Synchronous Flow Line

The MATLAB program was run through the variability scenarios to build the classification and regression data mining models for each test set and iterative training step w. Regression machine learning models were used to derive the formula while classification models were applied for comparison reasons along with the empirical formulas from the literature. Table 6.11 shows the score of the average and coefficient of variation of MAPE of each model, $\mu Score_m$ and $cScore_m$, for the iterative steps of the data set(s) used for training $\mathbf{S}_x$ and scoring criteria described in Section 4.5.3.2.1. Bold numbers indicate the methods with highest $\mu Score$ which are shortlisted based on top $cScore$ (shown in bold with light grey shading) for the iterative steps of train set. The elected method and train set (bold and dark grey shading) were chosen from the shortlisted methods of each individual test set as described in Section 4.5.3.2.1. The MAPE of the individual test data set e_{ma} within the training iterative step is shown in Appendix G (P. A-14).

In terms of the data set $\mathbf{D}_x, x \in \{1,2,3,4\}$ used as a training data set $\mathbf{S}_x$ for the data mining models, the data sets $\mathbf{D}_x, x = 1$ and $x = 2$, which include variability scenarios designed based on the number of factorial changes per experiment, performed poorly as training sets for the regression models. On the other hand, training the regression model with the full factorial DOE-based data sets of the same variability range of μ and c as $\mathbf{D}_x$, $x = 3$ and $x = 4$ gave better results in accuracy error presented by $\mu Score$ and stability across the test sets as measured by $cScore$ of the training set $\mathbf{R}_w$. As a result, for modelling of asynchronous flow lines, full factorial DOE only was used to define the variability scenarios. Furthermore, the data set $\mathbf{D}_x$, $x = 3$ performed better as a training set than the scaled-up data set $\mathbf{D}_x, x = 4$.

As expected testing with the same data sets used for training, i.e., supporting test set $\mathbf{T}_u$, normally gave less accuracy errors when compared with the main test set $\mathbf{T}_o$. However, for some experiments, the error was higher with the supporting test set $\mathbf{T}_u$ especially with the larger size data sets (Table G.3 and G.4) which suggests over-fitting of the data mining model. Hence, it is equally important to test the model prediction accuracy and stability over the full range of data sets including the training data sets.

Finally, for the modelling method DOF, The best performing comparison model across all iterative training steps w is the Blumenfeld (1990) formula followed by the Feedforward Neural Network. These models while trained with the data set $\mathbf{D}_x, x = 3$ scored 11 and 9 in terms of μ_e, respectively, and 1 in c_e.

As for the regression models, linear regression models based on robust fitting and regularisation algorithms to adjust the squared errors, which both exclude the interactions and higher orders of the selected predictor terms, performed the least in comparison to the stepwise regression. Additionally, regression models with bounded steps appear to be more accurate than unbounded regression, though in some cases both produce the same model. Purequadratic regression models are an exception since in this model only the bounded regression model has more limitations on covariate terms than unbounded. The regression steps for bounded purequadratic regression model include linear and squared terms only while unbounded regression model can add terms of a lower model form outside its boundary, e.g., multiplication of linear terms. In general, from the operational point of view of the stepwise regression, forward iteration in bounded regression performed better than backward iteration in unbounded regression when both have the same DOF in the covariate terms.

The polynomial stepwise regression model with bounded steps trained with data set $\mathbf{D}_x, x = 3$ gave the minimal prediction percentage error μ_e of 0.2% with a stability over the range of test data sets, i.e., $c_e = 0.11$. This model surpasses the performance of Blumenfeld (1990) formula (Figure 6.8-11) which gave for the same test data sets an average and coefficient of variation of MAPE, μ_e and c_e, of 2.63% and 0.74 respectively.

The multiple regression model of the throughput rate of synchronous flow line $\mathbf{TR}^{sync}$ for this regression model is given by:

$$\mathbf{TR}^{sync} = \mathbf{V}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{6.3}$$

where $\mathbf{V}$ is a $j \times 19$ matrix, $\boldsymbol{\beta}$ is a 19-dimensional vector and ε is a j-dimensional vector (see Table 6.10).

As shown in Equation 6.3 (Table 6.10), the stepwise regression model excluded predictor terms e^c, N and depended only on c, μ^{-1} and N^{-1} to formulate the relationship between variability within the synchronous flow line and TR.

Figure 6.12 shows the predicted against residuals plot using the optimal model for each test data set which suggests that the residuals are significantly small and stable over the full range of test data sets. This was also verified with the histogram plots shown in Appendix I (P. A-153). The residuals are generated primarily by an increase in c followed by a decline in μ and then shorter N (Figure I-5-7 in Appendix).

Table 6.10: Regression Model of the Throughput Rate for Synchronous Flow Lines

$$\mathbf{TR}^{sync} = \begin{pmatrix} TR_1^{sync} \\ TR_2^{sync} \\ \cdot \\ \cdot \\ \cdot \\ TR_j^{sync} \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} c_1 & N_1^{-1} & \mu_1^{-1} & c_1 \cdot N_1^{-1} & c_1 \cdot \mu_1^{-1} & N_1^{-1} \cdot \mu_1^{-1} & {c_1}^2 & c_1 \cdot N_1^{-1} \cdot \mu_1^{-1} & {c_1}^2 \cdot N_1^{-1} & {c_1}^2 \cdot \mu_1^{-1} & {c_1}^3 & {c_1}^2 \cdot N_1^{-1} \cdot \mu_1^{-1} & {c_1}^3 \cdot N_1^{-1} & {c_1}^3 \cdot \mu_1^{-1} & {c_1}^4 & {c_1}^3 \cdot N_1^{-1} \cdot \mu_1^{-1} & {c_1}^4 \cdot N_1^{-1} & {c_1}^4 \cdot \mu_1^{-1} & {c_1}^5 \\ c_2 & N_2^{-1} & \mu_2^{-1} & c_2 \cdot N_2^{-1} & c_2 \cdot \mu_2^{-1} & N_2^{-1} \cdot \mu_2^{-1} & {c_2}^2 & c_2 \cdot N_2^{-1} \cdot \mu_2^{-1} & {c_2}^2 \cdot N_2^{-1} & {c_2}^2 \cdot \mu_2^{-1} & {c_2}^3 & {c_2}^2 \cdot N_2^{-1} \cdot \mu_2^{-1} & {c_2}^3 \cdot N_2^{-1} & {c_2}^3 \cdot \mu_2^{-1} & {c_2}^4 & {c_2}^3 \cdot N_2^{-1} \cdot \mu_2^{-1} & {c_2}^4 \cdot N_2^{-1} & {c_2}^4 \cdot \mu_2^{-1} & {c_2}^5 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ c_j & N_j^{-1} & \mu_j^{-1} & c_j \cdot N_j^{-1} & c_j \cdot \mu_j^{-1} & N_j^{-1} \cdot \mu_j^{-1} & {c_j}^2 & c_j \cdot N_j^{-1} \cdot \mu_j^{-1} & {c_j}^2 \cdot N_j^{-1} & {c_j}^2 \cdot \mu_j^{-1} & {c_j}^3 & {c_j}^2 \cdot N_j^{-1} \cdot \mu_j^{-1} & {c_j}^3 \cdot N_j^{-1} & {c_j}^3 \cdot \mu_j^{-1} & {c_j}^4 & {c_j}^3 \cdot N_j^{-1} \cdot \mu_j^{-1} & {c_j}^4 \cdot N_j^{-1} & {c_j}^4 \cdot \mu_j^{-1} & {c_j}^5 \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} -0.0050908 \\ -0.00085394 \\ 0.99877 \\ 0.02319 \\ -1.2599 \\ 0.0072199 \\ 0.039156 \\ 1.3646 \\ -0.1328 \\ 1.2942 \\ -0.099294 \\ -1.7183 \\ 0.22185 \\ -0.84452 \\ 0.1037 \\ 0.72442 \\ -0.11021 \\ 0.24074 \\ -0.038618 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = 0.00014206 \begin{pmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}$$

Table 6.11: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Synchronous Flow Lines including Elected Training Set (Sn) and Method (m)

								Train Set R_w at x=1		Train Set R_w at x=2		Train Set R_w at x=3		Train Set R_w at x=4	
								I-II-S-1		*I-II-S-2*		*I-II-S-3*		*I-II-S-4*	
	Method				**Model Specifications**			**Score**		**Score**		**Score**		**Score**	
Number (m)	**Class I**	**Class II**	**Class III**	**Class IV**	**Data Mining Type**	**Learning Technique**	**Purpose**	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	1	0	1	0	1	1	2	1
2		Bootstrap Aggregating						0	1	0	0	1	1	1	1
3	Neural Network	Feedforward						0	1	0	0	9	1	0	0
4	Blumenfeld (1990)				Current Formula (Literature)	N/A		11	1	11	1	11	1	11	1
5	**Multiple Linear Regression**	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	0	0	0	0	2	1	2	1
6			Andrews					0	0	0	0	2	1	2	1
7			Cauchy M-estimators by Moore					**1**	0	0	0	2	1	2	1
8			Fair by Rey					**1**	0	0	1	2	1	2	2
9			Huber					**1**	0	0	1	2	1	2	2
10			Logistic Regression					**1**	0	0	1	2	1	2	2
11			Hinch and Talwar					0	0	0	0	2	1	2	2
12			Holland and Welsch					**1**	0	0	0	2	1	2	1
13		Regularisation	Lasso					**1**	**1**	**1**	**1**	2	1	2	1
14			Ridge Regression					**1**	0	**1**	**1**	2	1	2	1
15			Elastic Nets					**1**	**1**	**1**	**1**	2	1	2	1
16		**Stepwise**	Interaction	Bounded Steps				**1**	0	**1**	0	13	1	11	1
17				Unbounded Steps				**1**	0	**1**	0	13	1	11	1
18			Purequadratic	Bounded Steps				**1**	0	**1**	1	2	1	2	1
19				Unbounded Steps				**1**	0	**1**	0	13	1	11	1
20			Quadratic	Bounded Steps				**1**	0	**1**	0	13	1	11	1
21				Unbounded Steps				**1**	0	**1**	0	13	1	11	1
22			**Polynomial**	**Bounded Steps**				**1**	0	**1**	0	**20**	**2**	**20**	**1**
23				Unbounded Steps				**1**	0	**1**	0	13	1	13	1

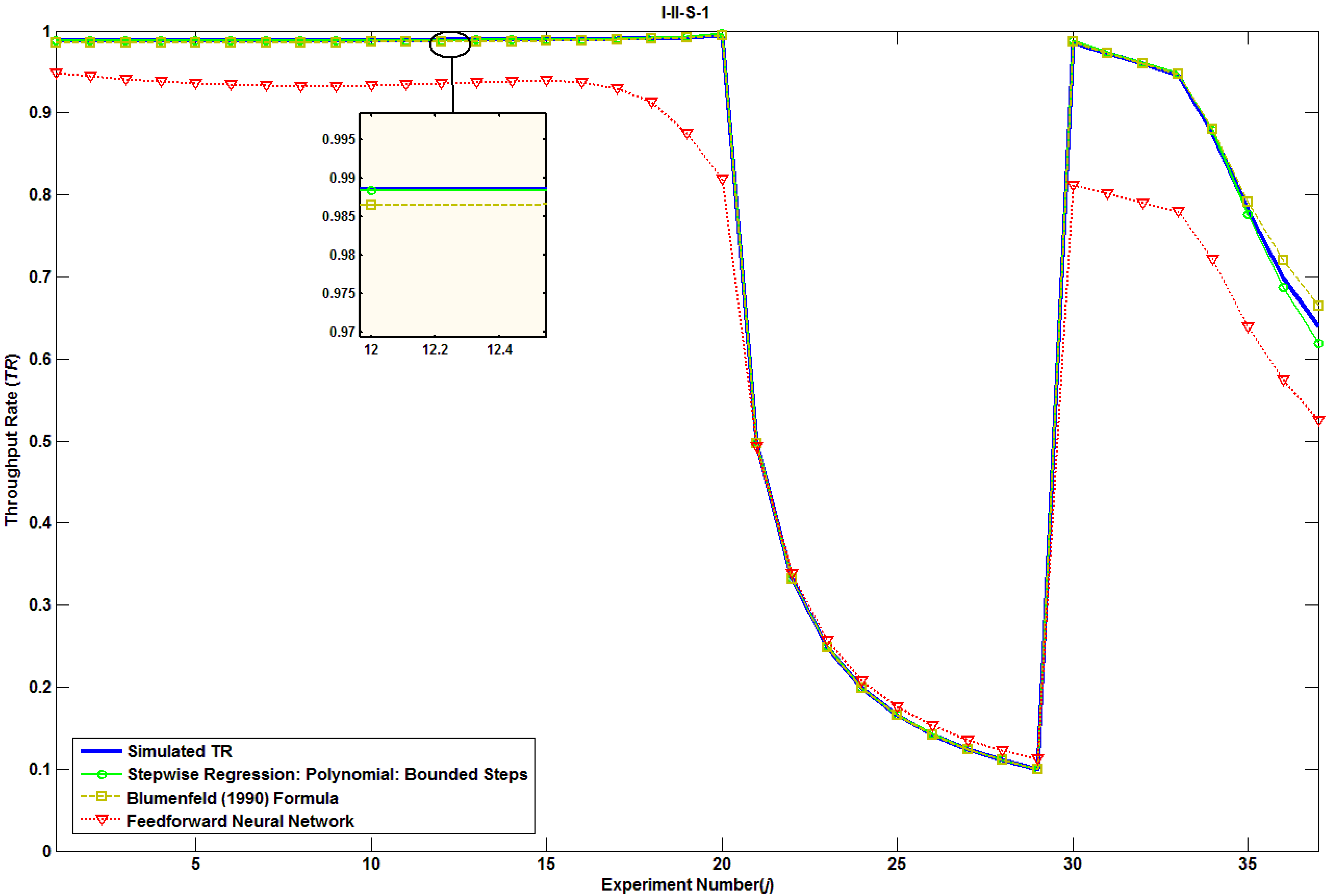


Figure 6.8: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-S-1$

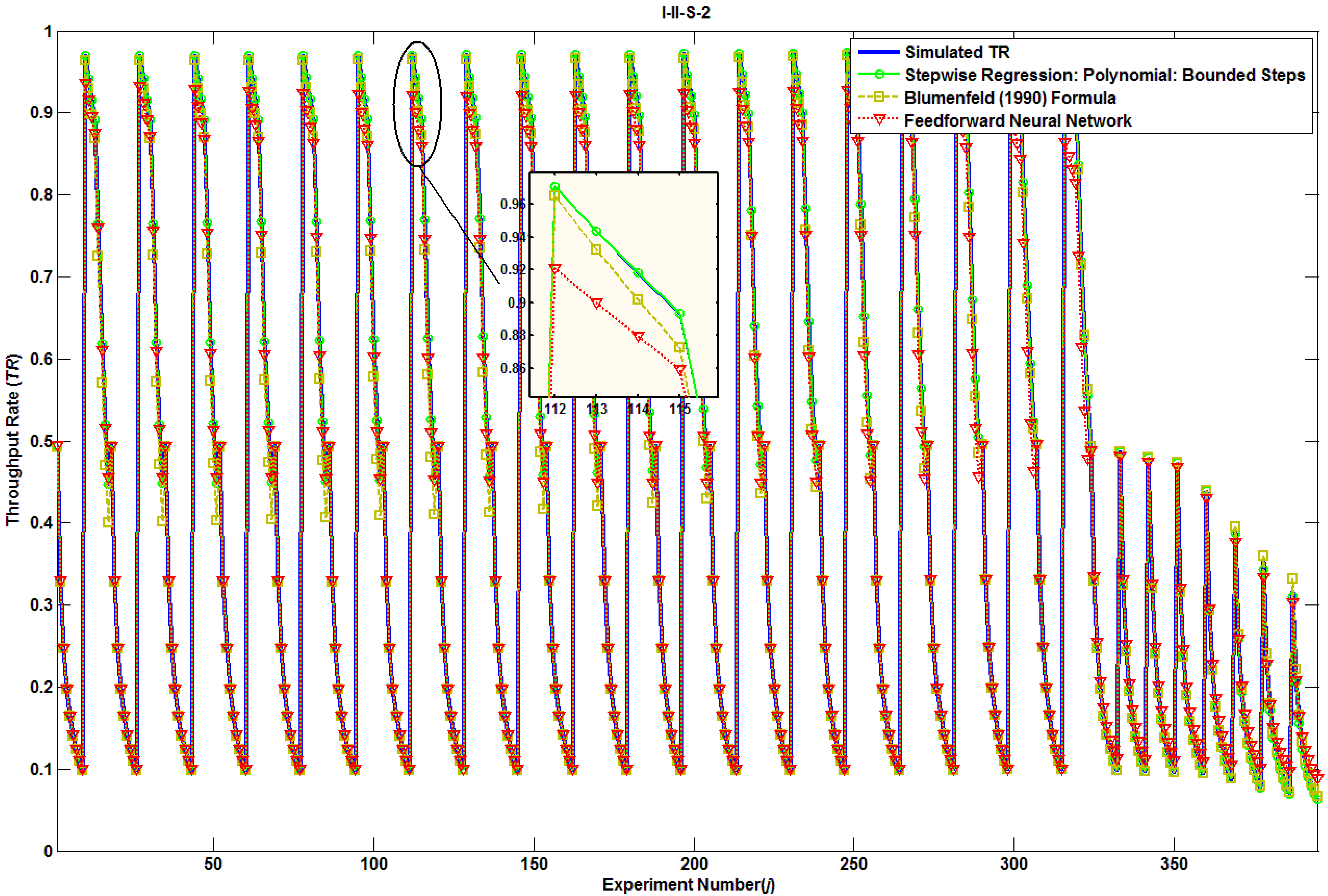


Figure 6.9: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-S-2$

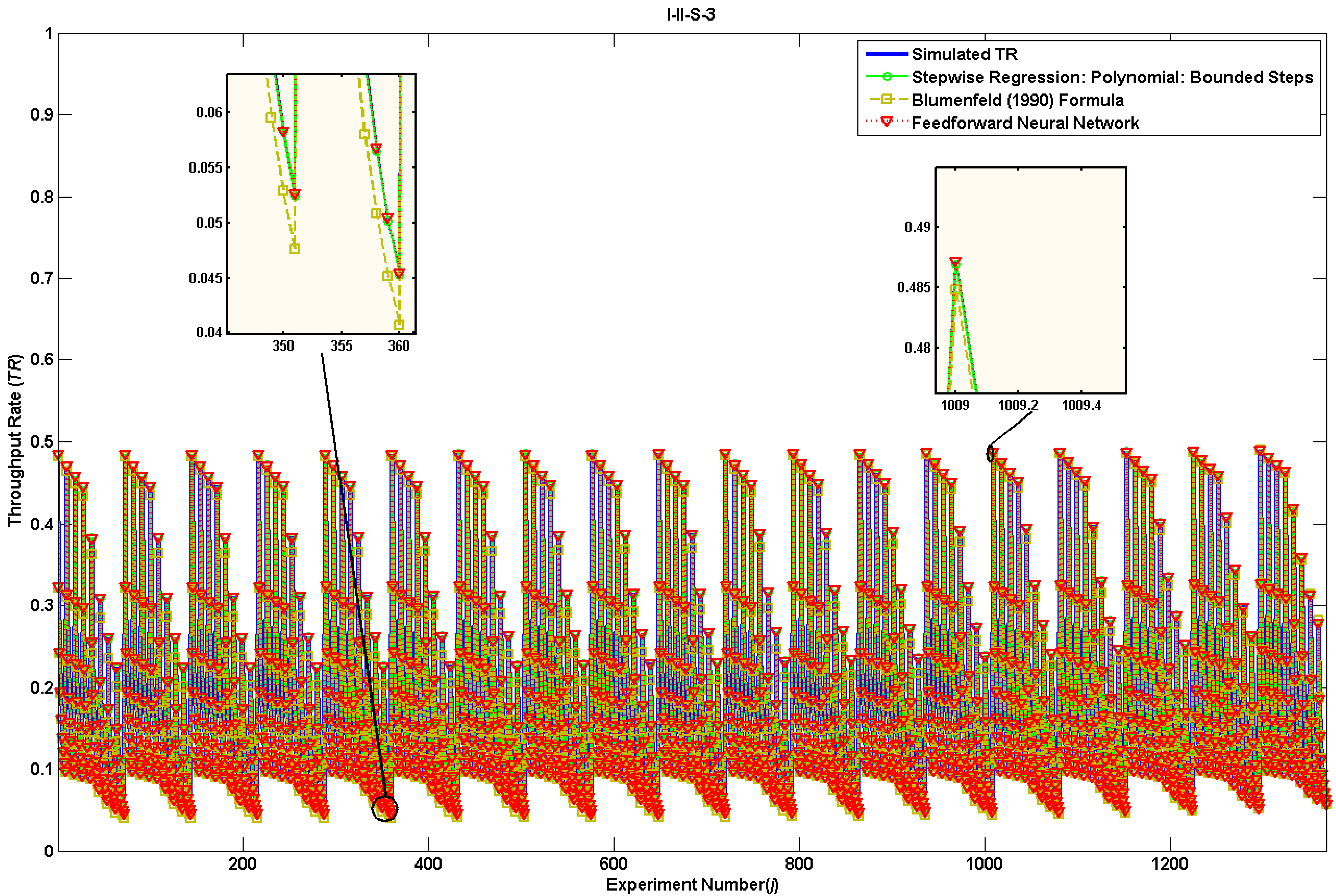


Figure 6.10: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-S-3$

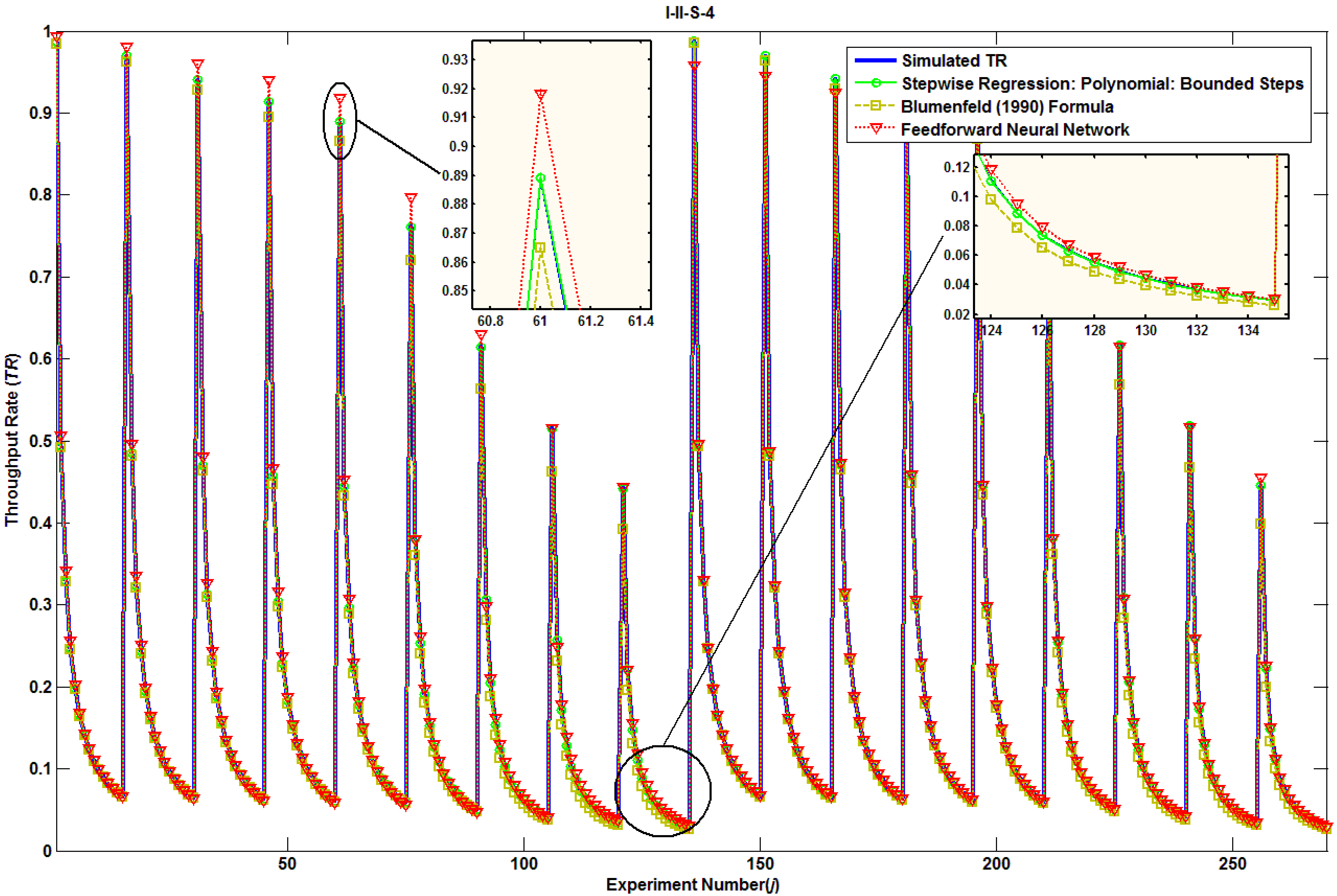


Figure 6.11: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-S-4$

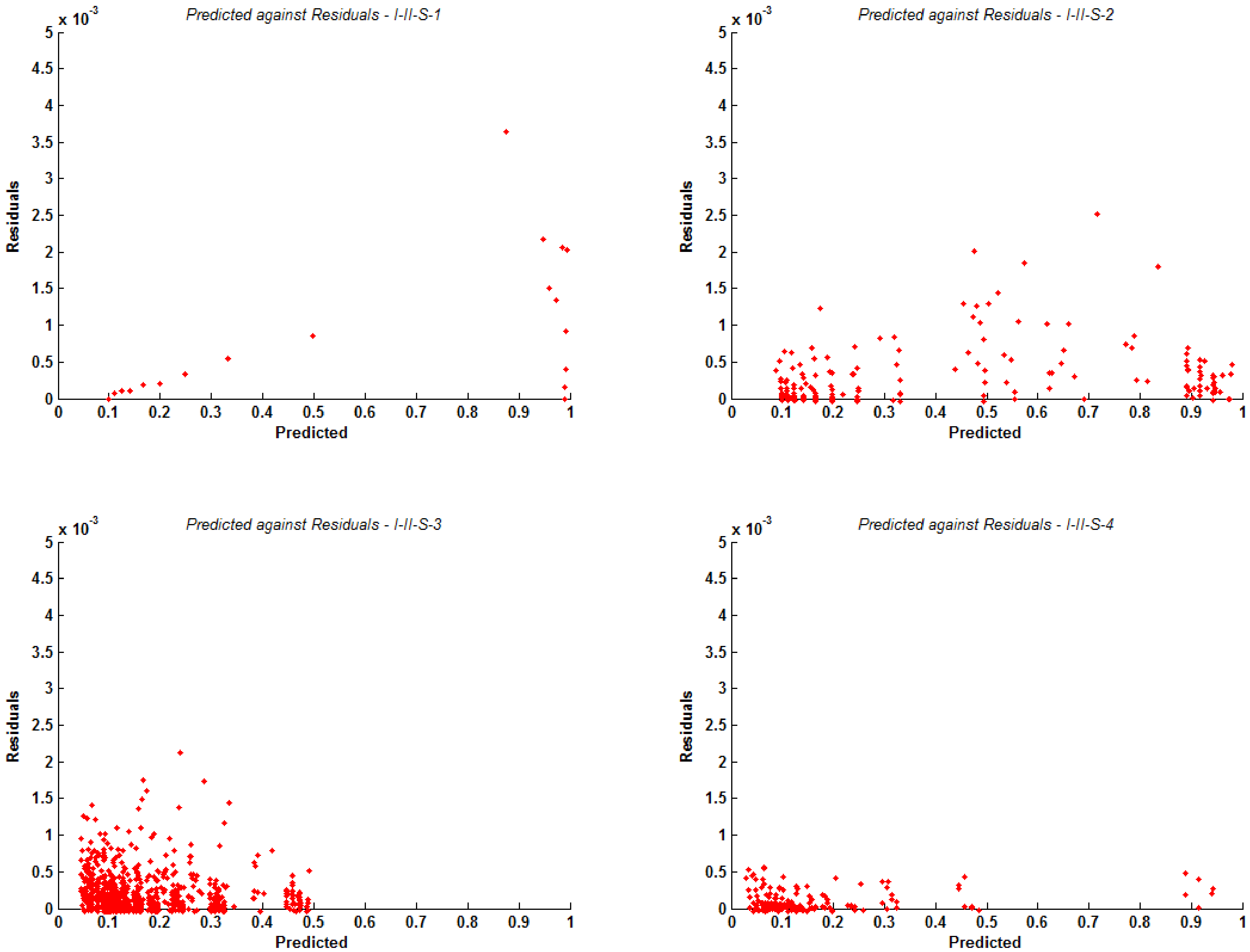


Figure 6.12: Predicted against Residuals Plots using the Optimal Regression Model for Test Data Sets $I-II-S-1$ **to** $I-II-S-4$

6.4.2 Asynchronous Flow Line

The data mining model for asynchronous flow line was built by running through the same DOFs as reported in Section 6.4.1 in addition to the two DOFs which were not applicable to the synchronous flow lines case, i.e., training set size and the addition of supporting predictor terms.

6.4.2.1 Training Set Size

The training set size represents the number of data sets $\mathbf{D}_x$ that are included as training data sets $\mathbf{S}_n$ within the training set $\mathbf{R}_w$. The size here can vary from a single, i.e., $n = 1$ to the total number of data sets $n = x = 8$. Decision on the number of data sets to consider was based on the improvements achieved from each iterative step within this DOF. When no further improvement could be achieved, the number of the best performing training set was maintained.

6.4.2.1.1 Single Data Set $(n = 1)$

The classification and regression data mining models were built using the developed MATLAB program for each test set and iterative training step w with a single data set within the training set, i.e., $n = 1$. Table 6.12 shows the score of the average and coefficient of variation of MAPE of each model, $\mu Score_m$ and $cScore_m$, for the iterative steps of the data sets used for training $\mathbf{S}_n$. The blank cells indicate a hardware or software limitation to perform modelling for these particular cases.

Among all data sets $\mathbf{D}_x, x \in \{1,2,...,8\}$ looped within the training data set $\mathbf{S}_n$ for the data mining models, the data set $\mathbf{D}_x$, $x = 8$ performed the best with the highest $\mu Score$ and $cScore$ .Three regression models with the same score were elected in this case; interaction, quadratic and polynomial regression, all with bounded steps.

In terms of methods and models performance, the deterministic throughput rate formula performed poorly for all test sets with $\mu Score$ and $cScore$ of 2 and 1, respectively, which suggests that variability was well introduced within the data sets. The robust fitting and regularisation algorithms of regression models were still generating high errors. Furthermore, bounded purequadratic regression without the multiplication of terms remains the worst among all stepwise regression along with polynomial regression with unbounded steps, i.e., backward iteration.

In general, Li and Meerkov (2009) formula performed the best among comparison and regression models including the elected three models, hence, double training data sets were examined to improve regression models prediction performance.

6.4.2.1.2 Double Data Sets $(n = 2)$

The best performing data set $\mathbf{D}_x, x = 8$ as a training data set at index $n = 1$ was kept the same while the index $n = 2$ was examined with the remaining seven data sets (Table 6.13).

Data set $\mathbf{D}_x, x = 3$ with the same three regression models; interaction, quadratic and polynomial regression, was the best performer. The statements in Section 6.4.2.1.1 regarding the performance of other models still hold true for this iterative step. The $\mu Score$ now increased from 9 to 11 with the added training data set $\mathbf{D}_x, x = 3$. However, it is still below $\mu Score$ for Li and Meerkov (2009) formula, i.e., 13.

6.4.2.1.3 Triple Data Sets $(n = 3)$

An additional data set was added to the training set at index $n = 3$ and looped over the data sets $\mathbf{D}_x$, $x \in \{1,2,3,4,5,6,7\}$ (Table 6.14). Polynomial regression model with bounded steps using the training set $\mathbf{D}_x, x = 4$ improved the $cScore$ to 2 which is higher than Li and Meerkov (2009) formula. However, the $\mu Score$ of Li and Meerkov (2009) formula still surpasses the polynomial model.

It is worth noting that addition of the wrong data set to the training set can reduce the performance such as in the case of $\mathbf{D}_x, x = 6$ and $x = 7$.

6.4.2.1.4 Quadruple Data Sets $(n = 4)$

Four indexed data sets $n \in \{1,2,3,4\}$ were included in the training set in this case (Table 6.15). In the indexes $n \in \{1,2,3,4\}$, the elected training data sets were added and the remaining data sets were examined at index $n = 4$.

No improvements were evident in this step, hence, the triple training data sets as elected in step 6.4.2.1.3 were used in the next DOF iterative steps, i.e., addition of supporting predictors.

The full results for the individual test set within this training iterative step are shown in Appendix H (P. A-23).

6.4.2.2 Supporting Predictor Terms

The training set was set from this step forward to always include the elected triple training data sets as outlined in Section 6.4.2.1 for the DOF related to the supporting predictors to commence. In this step, the supporting predictor terms were iteratively added to the training of machine learning models. The nine supporting predictors $p_j, j \in \{1,2,...,8\}$ are $e^{\mu_{\min}^{-1}}$, μ, μ^{-1}, $\log\mu$, $e^{\mu^{-1}}$, c, c^{-1} and $\log c$.

6.4.2.2.1 Single Supporting Predictor Term (i=1)

As shown in Table 6.17, in terms of regression models performance, robust fitting performance improved with the addition of a single supporting predictor term $p_j, j \in \{1,2,...,8\}$ to become comparable to that of the stepwise regression, excluding polynomial bounded regression, while regularisation of squared errors remained poor performer with no difference between the three algorithms with different penalties.

The supporting predictor $e^{\mu^{-1}}$ improved the prediction accuracy such that $\mu Score$ with the bounded steps polynomial regression model reached the same as the non-standalone Li and Meerkov (2009) formula while maintaining the $cScore$ at its higher value, i.e., 2. The best performing standalone regression model gave μ_e and c_e of 2% and 0.19 against 2% and 0.45 for Li and Meerkov (2009) formula.

Figure 6.13-20 show the predicted throughput rate of the optimal model against the simulated throughput rate while comparing it with Li and Meerkov (2009) formula and the best performing classification machine learning model for each individual test set. Due to the large size of data set $\mathbf{D}_x, x = 2$, only a sample of 1,100 variability scenarios are shown in Figure 6.14.

Table 6.12: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Single Data Set

Number (m)	Method				Model Specifications			Train Set R_w at n=1,x=1 *I-II-A-1* Score		Train Set R_w at n=1,x=2 *I-II-A-2* Score		Train Set R_w at n=1,x=3 *I-II-A-3* Score		Train Set R_w at n=1,x=4 *I-II-A-4* Score	
	Class I	Class II	Class III	Class IV	Data Mining Type	Learning Technique	Purpose	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	0	0	0	0	0	0	0	0
2		Bootstrap Aggregating						0	0	0	0	0	0	0	0
3	Neural Network	Feedforward						0	0			0	0	0	0
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A		13	1	13	1	13	1	13	1
5	$1/\mu_{\max}$							2	1	2	1	2	1	2	1
6	Multiple Linear Regression	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	0	0	**1**	0	0	0	2	0
7			Andrews					0	0	**1**	0	0	0	2	0
8			Cauchy M-estimators by Moore					0	0	**1**	0	0	0	2	0
9			Fair by Rey					0	0	0	0	0	0	1	0
10			Huber					0	0	0	0	0	0	1	0
11			Logistic Regression					0	0	0	0	0	0	1	0
12			Hinch and Talwar					0	0	**1**	0	0	0	2	0
13			Holland and Welsch					0	0	**1**	0	0	0	2	0
14		Regularisation	Lasso					0	0			0	0	2	0
15			Ridge Regression					0	0			0	0	2	0
16			Elastic Nets					0	0			0	0	2	0
17		Stepwise	Interaction	Bounded Steps				0	0	**1**	0	**1**	0	2	0
18				Unbounded Steps				0	0	**1**	0	**1**	0	0	0
19			Purequadratic	Bounded Steps				0	0	0	0	0	0	2	0
20				Unbounded Steps				0	0	**1**	0	**1**	0	**3**	**1**
21			Quadratic	Bounded Steps				0	0	**1**	0	**1**	0	2	0
22				Unbounded Steps				0	0	**1**	0	**1**	0	0	0
23			Polynomial	Bounded Steps				0	0	**1**	0	**1**	0	2	0
24				Unbounded Steps				0	0			**1**	0	0	0

Table 6.12: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Single Data Set (*cont.*)

								Train Set R_w at n=1,x=5		Train Set R_w at n=1,x=6		Train Set R_w at n=1,x=7		Train Set R_w at n=1,x=8	
								I-II-A-5		*I-II-A-6*		*I-II-A-7*		*I-II-A-8*	
	Method				Model Specifications			Score		Score		Score		Score	
Number (m)	Class I	Class II	Class III	Class IV	Data Mining Type	Learning Technique	Purpose	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	1	0	0	0	1	0	1	1
2		Bootstrap Aggregating						0	0	0	0	1	0	1	1
3	Neural Network	Feedforward						1	0	0	0	1	0	1	1
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A		13	1	13	1	13	1	13	1
5	$1/\mu_{\max}$							2	1	2	1	2	1	2	1
6	Multiple Linear Regression	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	0	0	1	0	2	0	4	1
7			Andrews					0	0	1	0	2	0	4	1
8			Cauchy M-estimators by Moore					0	0	1	0	2	0	4	1
9			Fair by Rey					1	0	1	0	2	0	3	1
10			Huber					0	0	1	0	2	0	3	1
11			Logistic Regression					0	0	1	0	2	0	3	1
12			Hinch and Talwar					0	0	1	0	2	0	4	1
13			Holland and Welsch					0	0	1	0	2	0	4	1
14		Regularisation	Lasso					2	0	**2**	0	2	0	1	1
15			Ridge Regression					2	0	**2**	0	2	0	1	1
16			Elastic Nets					2	0	**2**	0	2	0	1	1
17		**Stepwise**	**Interaction**	**Bounded Steps**				**7**	**1**	**2**	0	**7**	**1**	**9**	**1**
18				Unbounded Steps				**7**	**1**	1	0	2	0	2	1
19			Purequadratic	Bounded Steps				1	1	**2**	0	2	0	2	1
20				Unbounded Steps				**7**	**1**	0	0	2	0	2	1
21			**Quadratic**	**Bounded Steps**				**7**	**1**	**2**	0	**7**	**1**	**9**	**1**
22				Unbounded Steps				**7**	**1**	1	0	2	0	2	1
23			**Polynomial**	**Bounded Steps**				**7**	**1**	**2**	0	**7**	**1**	**9**	**1**
24				Unbounded Steps				2	1	**2**	0	2	0	2	1

Table 6.13: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Double Data Sets

								Train Set R_w at n=2,x=1		Train Set R_w at n=2,x=2		Train Set R_w at n=2,x=3		Train Set R_w at n=2,x=4	
								I-II-A-1		*I-II-A-2*		*I-II-A-3*		*I-II-A-4*	
	Method				**Model Specifications**			**Score**		**Score**		**Score**		**Score**	
Number (m)	**Class I**	**Class II**	**Class III**	**Class IV**	**Data Mining Type**	**Learning Technique**	**Purpose**	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	0	1	0	0	0	0	1	1
2		Bootstrap Aggregating						0	0			0	0	1	1
3	Neural Network	Feedforward						0	0			1	0	4	0
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A		13	1	13	1	13	1	13	1
5	$\frac{1}{\mu_{max}}$							2	1	2	1	2	1	2	1
6	**Multiple Linear Regression**	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	1	0	1	0	2	1	4	1
7			Andrews					1	0	1	0	2	1	4	1
8			Cauchy M-estimators by Moore					2	0	1	0	1	0	4	1
9			Fair by Rey					1	0	0	0	1	0	3	1
10			Huber					1	0	0	0	1	0	4	1
11			Logistic Regression					1	0	0	0	1	0	4	1
12			Hinch and Talwar					4	1	1	0	1	0	4	1
13			Holland and Welsch					1	0	1	0	2	1	4	1
14		Regularisation	Lasso					1	0			1	0	2	0
15			Ridge Regression					1	0			1	0	2	0
16			Elastic Nets					1	0			1	0	2	0
17		**Stepwise**	**Interaction**	**Bounded Steps**				2	0	**2**	0	**11**	**1**	2	0
18				Unbounded Steps				2	0	**2**	0	2	0	5	1
19			Purequadratic	Bounded Steps				1	0	0	0	0	0	2	0
20				Unbounded Steps				1	0	**2**	0	1	0	**6**	**1**
21			**Quadratic**	**Bounded Steps**				2	0	**2**	0	**11**	**1**	**6**	**1**
22				Unbounded Steps				2	0	**2**	0	1	0	**6**	**1**
23			**Polynomial**	**Bounded Steps**				**9**	**2**	**2**	0	**11**	**1**	2	0
24				Unbounded Steps				3	0			0	0	0	0

Table 6.13: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Double Data Sets (*cont.*)

	Method				Model Specifications			Train Set R_w at n=2,x=5 *I-II-A-5* Score		Train Set R_w at n=2,x=6 *I-II-A-6* Score		Train Set R_w at n=2,x=7 *I-II-A-7* Score	
Number (m)	**Class I**	**Class II**	**Class III**	**Class IV**	**Data Mining Type**	**Learning Technique**	**Purpose**	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	1	1	1	1	1	1
2	Decision Tree	Bootstrap Aggregating			Supervised Machine Learning	Classification	Comparison	1	1	1	1	1	1
3	Neural Network	Feedforward			Supervised Machine Learning	Classification	Comparison	1	1	1	1	1	0
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A	Comparison	13	1	13	1	13	1
5	$1/\mu_{\max}$				Current Formula (Literature)	N/A	Comparison	2	1	2	1	2	1
6	Multiple Linear Regression	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	4	1	4	1	2	1
7	Multiple Linear Regression	Robust	Andrews		Supervised Machine Learning	Regression	Formula Derivation	4	1	4	1	2	1
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		Supervised Machine Learning	Regression	Formula Derivation	4	1	4	1	2	1
9	Multiple Linear Regression	Robust	Fair by Rey		Supervised Machine Learning	Regression	Formula Derivation	3	1	4	1	2	1
10	Multiple Linear Regression	Robust	Huber		Supervised Machine Learning	Regression	Formula Derivation	3	1	4	1	2	1
11	Multiple Linear Regression	Robust	Logistic Regression		Supervised Machine Learning	Regression	Formula Derivation	3	1	4	1	2	1
12	Multiple Linear Regression	Robust	Hinch and Talwar		Supervised Machine Learning	Regression	Formula Derivation	4	1	4	1	2	1
13	Multiple Linear Regression	Robust	Holland and Welsch		Supervised Machine Learning	Regression	Formula Derivation	4	1	4	1	2	1
14	Multiple Linear Regression	Regularisation	Lasso		Supervised Machine Learning	Regression	Formula Derivation	2	1	2	1	2	1
15	Multiple Linear Regression	Regularisation	Ridge Regression		Supervised Machine Learning	Regression	Formula Derivation	2	1	2	1	2	1
16	Multiple Linear Regression	Regularisation	Elastic Nets		Supervised Machine Learning	Regression	Formula Derivation	2	1	2	1	2	1
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	Supervised Machine Learning	Regression	Formula Derivation	**9**	**1**	**9**	**1**	**9**	**1**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	Supervised Machine Learning	Regression	Formula Derivation	**9**	**1**	**9**	**1**	**9**	**1**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	Supervised Machine Learning	Regression	Formula Derivation	2	1	0	0	2	1
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	Supervised Machine Learning	Regression	Formula Derivation	**9**	**1**	**9**	**1**	**9**	**1**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	Supervised Machine Learning	Regression	Formula Derivation	**9**	**1**	**9**	**1**	**9**	**1**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	Supervised Machine Learning	Regression	Formula Derivation	**9**	**1**	**9**	**1**	**9**	**1**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	Supervised Machine Learning	Regression	Formula Derivation	**9**	**1**	**9**	**1**	**9**	**1**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	Supervised Machine Learning	Regression	Formula Derivation	2	0	**9**	**1**	1	0

Table 6.14: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Triple Data Sets

Number (m)	Method: Class I	Class II	Class III	Class IV	Model Specifications: Data Mining Type	Learning Technique	Purpose	Train Set R_w at n=3,x=1 I-II-A-1 Score $\mu Score_m$	$cScore_m$	Train Set R_w at n=3,x=2 I-II-A-2 Score $\mu Score_m$	$cScore_m$	Train Set R_w at n=3,x=4 I-II-A-4 Score $\mu Score_m$	$cScore_m$	Train Set R_w at n=3,x=5 I-II-A-5 Score $\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	0	0	0	0	1	1	1	1
2		Bootstrap Aggregating						0	0			1	1	1	1
3	Neural Network	Feedforward						1	1			2	0	0	0
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A		13	1	13	1	13	1	13	1
5	$1/\mu_{\max}$							2	1	2	1	2	1	2	1
6	**Multiple Linear Regression**	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	1	0	1	0	4	1	3	1
7			Andrews					1	0	1	0	4	1	3	1
8			Cauchy M-estimators by Moore					1	0	1	0	4	1	4	1
9			Fair by Rey					1	0	0	0	2	1	2	1
10			Huber					1	0	0	0	3	1	2	1
11			Logistic Regression					1	0	0	0	3	1	2	1
12			Hinch and Talwar					1	0	1	0	4	1	2	1
13			Holland and Welsch					1	0	1	0	4	1	3	1
14		Regularisation	Lasso					1	0			2	0	2	0
15			Ridge Regression					1	0			2	0	2	0
16			Elastic Nets					1	0			2	0	2	0
17		**Stepwise**	Interaction	Bounded Steps				2	0	**2**	0	2	0	**11**	**1**
18				Unbounded Steps				2	0	**2**	**1**	5	1	**11**	**1**
19			Purequadratic	Bounded Steps				1	0	**0**	0	2	0	2	0
20				Unbounded Steps				2	0	**2**	0	5	1	9	1
21			Quadratic	Bounded Steps				2	0	**2**	0	3	0	**11**	**1**
22				Unbounded Steps				2	0	**2**	0	5	1	9	1
23			**Polynomial**	**Bounded Steps**				**9**	**1**	**2**	0	**11**	**2**	**11**	**1**
24				Unbounded Steps				2	1			1	0	1	0

Table 6.14: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Triple Data Sets *(cont.)*

								Train Set R_w at n=3,x=6 *I-II-A-6*		Train Set R_w at n=3,x=7 *I-II-A-7*	
	Method				**Model Specifications**			**Score**		**Score**	
Number (m)	**Class I**	**Class II**	**Class III**	**Class IV**	**Data Mining Type**	**Learning Technique**	**Purpose**	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	0	0	0	0
2		Bootstrap Aggregating						0	0	0	0
3	Neural Network	Feedforward						1	0	1	0
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A		13	1	13	1
5	$1/\mu_{max}$							2	1	2	1
6	Multiple Linear Regression	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	**4**	**1**	1	0
7			Andrews					**4**	**1**	1	0
8			Cauchy M-estimators by Moore					**4**	**1**	**2**	0
9			Fair by Rey					2	1	1	0
10			Huber					3	1	1	0
11			Logistic Regression					3	1	**2**	0
12			Hinch and Talwar					**4**	**1**	**2**	**1**
13			Holland and Welsch					**4**	**1**	**2**	0
14		Regularisation	Lasso					1	0	1	0
15			Ridge Regression					1	0	1	0
16			Elastic Nets					1	0	1	0
17		Stepwise	Interaction	Bounded Steps				2	0	**2**	0
18				Unbounded Steps				2	0	**2**	0
19			Purequadratic	Bounded Steps				1	0	1	0
20				Unbounded Steps				1	0	1	0
21			Quadratic	Bounded Steps				2	0	**2**	0
22				Unbounded Steps				1	0	1	0
23			Polynomial	Bounded Steps				0	0	0	0
24				Unbounded Steps				0	0	0	0

Table 6.15: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Quadruple Data Sets

	Method				Model Specifications			Train Set R_w at n=4,x=1 *I-II-A-1* Score		Train Set R_w at n=4,x=2 *I-II-A-2* Score		Train Set R_w at n=4,x=5 *I-II-A-5* Score		Train Set R_w at n=4,x=6 *I-II-A-6* Score	
Number (m)	**Class I**	**Class II**	**Class III**	**Class IV**	**Data Mining Type**	**Learning Technique**	**Purpose**	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	1	0	1	1	1	0	1	0
2		Bootstrap Aggregating						1	1			1	0	1	0
3	Neural Network	Feedforward						1	0			2	0	11	1
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A		13	1	13	1	13	1	13	1
5	$1/\mu_{max}$							2	1	2	1	2	1	2	1
6	Multiple Linear Regression	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	2	0	1	0	2	0	4	1
7			Andrews					2	0	1	0	2	0	4	1
8			Cauchy M-estimators by Moore					4	1	1	0	3	1	4	1
9			Fair by Rey					1	0	0	0	2	0	2	0
10			Huber					1	0	0	0	2	0	3	1
11			Logistic Regression					1	0	0	0	2	0	3	1
12			Hinch and Talwar					4	1	1	0	2	0	2	0
13			Holland and Welsch					2	0	1	0	3	1	4	1
14		Regularisation	Lasso					1	0			2	0	2	0
15			Ridge Regression					1	0			2	0	2	0
16			Elastic Nets					1	0			2	0	2	0
17		Stepwise	Interaction	Bounded Steps				3	1	1	0	1	0	5	1
18				Unbounded Steps				2	1	1	0	7	1	4	1
19			Purequadratic	Bounded Steps				1	0	0	0	2	0	2	0
20				Unbounded Steps				2	0	**2**	0	7	1	5	1
21			Quadratic	Bounded Steps				2	1	1	1	3	0	5	1
22				Unbounded Steps				2	0	**2**	**1**	7	1	5	1
23			Polynomial	Bounded Steps				**9**	**1**	2	0	**11**	**2**	**11**	**1**
24				Unbounded Steps				7	1			2	0	2	0

Table 6.15: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Quadruple Data Sets *(cont.)*

Number (m)	Method: Class I	Method: Class II	Method: Class III	Method: Class IV	Model Specifications: Data Mining Type	Model Specifications: Learning Technique	Model Specifications: Purpose	Train Set R_w at n=4,x=7 *I-II-A-7* Score: $\mu Score_m$	Train Set R_w at n=4,x=7 *I-II-A-7* Score: $cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	1	0
2	Decision Tree	Bootstrap Aggregating			Supervised Machine Learning	Classification	Comparison	1	0
3	Neural Network	Feedforward			Supervised Machine Learning	Classification	Comparison	9	1
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A	Comparison	13	1
5	$1/\mu_{\max}$				Current Formula (Literature)	N/A	Comparison	2	1
6	Multiple Linear Regression	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	2	0
7	Multiple Linear Regression	Robust	Andrews		Supervised Machine Learning	Regression	Formula Derivation	2	0
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		Supervised Machine Learning	Regression	Formula Derivation	2	0
9	Multiple Linear Regression	Robust	Fair by Rey		Supervised Machine Learning	Regression	Formula Derivation	2	0
10	Multiple Linear Regression	Robust	Huber		Supervised Machine Learning	Regression	Formula Derivation	2	0
11	Multiple Linear Regression	Robust	Logistic Regression		Supervised Machine Learning	Regression	Formula Derivation	2	0
12	Multiple Linear Regression	Robust	Hinch and Talwar		Supervised Machine Learning	Regression	Formula Derivation	2	0
13	Multiple Linear Regression	Robust	Holland and Welsch		Supervised Machine Learning	Regression	Formula Derivation	2	0
14	Multiple Linear Regression	Regularisation	Lasso		Supervised Machine Learning	Regression	Formula Derivation	2	0
15	Multiple Linear Regression	Regularisation	Ridge Regression		Supervised Machine Learning	Regression	Formula Derivation	2	0
16	Multiple Linear Regression	Regularisation	Elastic Nets		Supervised Machine Learning	Regression	Formula Derivation	2	0
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	Supervised Machine Learning	Regression	Formula Derivation	4	1
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	Supervised Machine Learning	Regression	Formula Derivation	3	1
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	Supervised Machine Learning	Regression	Formula Derivation	2	0
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	Supervised Machine Learning	Regression	Formula Derivation	4	1
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	Supervised Machine Learning	Regression	Formula Derivation	5	1
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	Supervised Machine Learning	Regression	Formula Derivation	4	1
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	Supervised Machine Learning	Regression	Formula Derivation	**11**	**1**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	Supervised Machine Learning	Regression	Formula Derivation	2	0

The multiple regression model of the throughput rate of asynchronous flow line $\mathbf{TR}^{\text{async}}$ for this polynomial regression model is expressed as:

$$\mathbf{TR}^{\text{async}} = \mathbf{V}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{6.4}$$

where $\mathbf{V}$ is a $j \times 15$ matrix, $\boldsymbol{\beta}$ is a 15-dimensional vector and ε is a j-dimensional vector (see Table 10). ${\mu_{\max}}^{-1}$, $e^{c_{av}}$ and N^{-1}

The polynomial regression model (Equation 6.4 and Table 6.16) shows that the stepwise model excluded the following main predictor terms c_{av} and N while used ${\mu_{\max}}^{-1}$, $e^{c_{av}}$ and N^{-1} along with the supporting predictor $e^{\mu^{-1}}$ to formulate the relationship between process variability within the asynchronous flow line and TR.

The predicted against residuals plot using the optimal model for each test data set are shown in Figure 6.21. The residuals here are minor and stable, to a large extent, across the full range of test data sets which agrees with the histogram plots as shown in Appendix J (P. A-161). It is still evident that the errors are proportionally related to c_{av} while the relationship between errors and $\mu_{\max}$ is not steady and due to the nature of the data sets, the relationship with N cannot be examined (Figure J-9-11 in Appendix).

6.4.2.2.2 Double Supporting Predictor Terms (i=2)

Addition of another supporting predictor term to the training data sets failed to show any improvement to the prediction performance as shown in Table 6.18. Hence, the additional supporting predictor terms were kept to $e^{\mu^{-1}}$ only.

The full results for the individual test set within this training iterative step are shown in Appendix H (P. A-23).

Table 6.16: Regression Model of the Throughput Rate for Asynchronous Flow Lines

$$\mathbf{TR}^{\mathrm{async}} = \begin{pmatrix} TR_1^{async} \\ TR_2^{async} \\ . \\ . \\ . \\ TR_j^{async} \end{pmatrix},$$

$$\mathbf{V} = \left(\begin{array}{ccccccccccccccc}
e_1^{c_{av}} & \mu_{\max,1}^{-1} & N_1^{-1} & e_1^{\mu^{-1}} & e_1^{c_{av}} \cdot \mu_{\max,1}^{-1} & e_1^{c_{av}} \cdot N_1^{-1} & e_1^{c_{av}} \cdot e_1^{\mu^{-1}} & \mu_{\max,1}^{-1} \cdot N_1^{-1} & \mu_{\max,1}^{-1} \cdot e_1^{\mu^{-1}} & N_1^{-1} \cdot e_1^{\mu^{-1}} & \mu_{\max,1}^{-2} & e_1^{\mu^{-2}} & e_1^{c_{av}} \cdot \mu_{\max,1}^{-1} \cdot N_1^{-1} & e_1^{c_{av}} \cdot N_1^{-1} \cdot e_1^{\mu^{-1}} & e_1^{c_{av}} \cdot \mu_{\max,1}^{-2} \\
e_2^{c_{av}} & \mu_{\max,2}^{-1} & N_2^{-1} & e_2^{\mu^{-1}} & e_2^{c_{av}} \mu_{\max,2}^{-1} & e_2^{c_{av}} \cdot N_2^{-1} & e_2^{c_{av}} \cdot e_2^{\mu^{-1}} & \mu_{\max,2}^{-1} \cdot N_2^{-1} & \mu_{\max,2}^{-1} \cdot e_2^{\mu^{-1}} & N_2^{-1} \cdot e_2^{\mu^{-1}} & \mu_{\max,2}^{-2} & e_2^{\mu^{-2}} & e_2^{c_{av}} \mu_{\max,2}^{-1} \cdot N_2^{-1} & e_2^{c_{av}} \cdot N_2^{-1} \cdot e_2^{\mu^{-1}} & e_2^{c_{av}} \mu_{\max,2}^{-2} \\
. & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\
e_j^{c_{av}} & \mu_{\max,j}^{-1} & N_j^{-1} & e_j^{\mu^{-1}} & e_j^{c_{av}} \cdot \mu_{\max,j}^{-1} & e_j^{c_{av}} \cdot N_j^{-1} & e_j^{c_{av}} \cdot e_j^{\mu^{-1}} & \mu_{\max,j}^{-1} \cdot N_j^{-1} & \mu_{\max,j}^{-1} \cdot e_j^{\mu^{-1}} & N_j^{-1} \cdot e_j^{\mu^{-1}} & \mu_{\max,j}^{-2} & e_j^{\mu^{-2}} & e_j^{c_{av}} \cdot \mu_{\max,j}^{-1} \cdot N_j^{-1} & e_j^{c_{av}} \cdot N_j^{-1} \cdot e_j^{\mu^{-1}} & e_j^{c_{av}} \cdot \mu_{\max,j}^{-2}
\end{array}\right),$$

$$\boldsymbol{\beta} = \begin{pmatrix} -0.074113 \\ 1.2413 \\ 0.16525 \\ 0.047208 \\ -0.38561 \\ -0.15128 \\ 0.074081 \\ -0.088294 \\ 0.11468 \\ -0.16522 \\ 0.081924 \\ -0.051851 \\ 0.12008 \\ 0.15121 \\ -0.141977 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\varepsilon} = 0.0046669 \begin{pmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \end{pmatrix}$$

Table 6.17: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Single Supp. Predictor

								Train Set R_w at n=3,i=1		Train Set R_w at n=3,i=2		Train Set R_w at n=3,i=3		Train Set R_w at n=3,i=4	
								$e^{\mu_{\min}^{-1}}$		μ		μ^{-1}		$\log \mu$	
	Method				Model Specifications			Score		Score		Score		Score	
Number (m)	Class I	Class II	Class III	Class IV	Data Mining Type	Learning Technique	Purpose	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	1	1	1	1	1	1	1	1
2		Bootstrap Aggregating						1	0	1	0	1	0	1	0
3	Neural Network	Feedforward						2	0	0	0	2	0	1	0
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A		13	1	13	1	13	1	13	1
5	$1/\mu_{\max}$							2	1	2	1	2	1	2	1
6	Multiple Linear Regression	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	4	1	4	1	4	1	4	1
7			Andrews					4	1	4	1	4	1	4	1
8			Cauchy M-estimators by Moore					4	1	4	1	4	1	4	1
9			Fair by Rey					2	1	2	1	2	1	2	1
10			Huber					3	1	2	1	3	1	2	1
11			Logistic Regression					3	1	2	1	3	1	2	1
12			Hinch and Talwar					4	1	4	1	4	1	4	1
13			Holland and Welsch					4	1	4	1	4	1	4	1
14		Regularisation	Lasso					2	0	2	0	2	0	2	0
15			Ridge Regression					2	0	2	0	2	0	2	0
16			Elastic Nets					2	0	2	0	2	0	2	0
17		Stepwise	Interaction	Bounded Steps				2	1	2	0	7	1	2	1
18				Unbounded Steps				5	1	2	1	7	1	2	0
19			Purequadratic	Bounded Steps				2	0	2	0	2	0	1	0
20				Unbounded Steps				**5**	**1**	2	1	7	1	1	0
21			Quadratic	Bounded Steps				3	0	3	1	2	0	5	1
22				Unbounded Steps				**5**	**1**	2	1	2	0	2	0
23			Polynomial	Bounded Steps				2	0	**11**	**2**	**13**	**1**	**9**	**1**
24				Unbounded Steps				1	1	0	0				

Table 6.17: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Single Supp. Predictor (*cont.*)

								Train Set R_w at n=3,i=5		Train Set R_w at n=3,i=6		Train Set R_w at n=3,i=7		Train Set R_w at n=3,i=8	
								Train Set R_w at n=3,i=1		Train Set R_w at n=3,i=2		Train Set R_w at n=3,i=3		Train Set R_w at n=3,i=4	
	Method				Model Specifications			Score		Score		Score		Score	
Number (m)	Class I	Class II	Class III	Class IV	Data Mining Type	Learning Technique	Purpose	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	1	1	1	1	1	1	1	1
2		Bootstrap Aggregating						1	0	1	1	1	1	1	1
3	Neural Network	Feedforward						7	1	2	1	1	0	1	0
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A		13	1	13	1	13	1	13	1
5	$1/\mu_{max}$							2	1	2	1	2	1	2	1
6	Multiple Linear Regression	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	4	1	1	1	1	1	1	1
7			Andrews					4	1	1	1	1	1	1	1
8			Cauchy M-estimators by Moore					4	1	1	1	1	1	1	1
9			Fair by Rey					2	1	1	1	1	1	1	1
10			Huber					3	1	1	1	1	1	1	1
11			Logistic Regression					3	1	1	1	1	1	1	1
12			Hinch and Talwar					4	1	1	1	1	1	1	1
13			Holland and Welsch					4	1	1	1	1	1	1	1
14		Regularisation	Lasso					2	0	0	0	0	0	0	0
15			Ridge Regression					2	0	0	0	0	0	0	0
16			Elastic Nets					2	0	0	0	0	0	0	0
17		**Stepwise**	Interaction	Bounded Steps				6	1	2	0	2	0	2	0
18				Unbounded Steps				7	1	0	0	0	0	0	0
19			Purequadratic	Bounded Steps				2	0	2	0	2	0	2	0
20				Unbounded Steps				6	1	0	1	0	1	0	1
21			Quadratic	Bounded Steps				6	1	3	0	3	0	3	0
22				Unbounded Steps				2	0	0	1	0	1	0	1
23			**Polynomial**	**Bounded Steps**				**13**	**2**	**11**	**2**	**11**	**2**	**11**	**2**
24				Unbounded Steps											

Number (m)	Method: Class I	Class II	Class III	Class IV	Model Specifications: Data Mining Type	Learning Technique	Purpose	$e^{\mu_{\min}^{-1}}$ Score: $\mu Score_m$	$cScore_m$	μ Score: $\mu Score_m$	$cScore_m$	μ^{-1} Score: $\mu Score_m$	$cScore_m$	$\log \mu$ Score: $\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	1	1	1	1	1	1	1	1
2		Bootstrap Aggregating						1	0	2	0	1	0	2	0
3	Neural Network	Feedforward						2	1	1	0	2	0	2	0
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A		13	1	13	1	13	1	13	1
5	$1/\mu_{\max}$							2	1	2	1	2	1	2	1
6	Multiple Linear Regression	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	4	1	4	1	4	1	4	1
7			Andrews					4	1	4	1	4	1	4	1
8			Cauchy M-estimators by Moore					4	1	4	1	4	1	4	1
9			Fair by Rey					2	1	2	1	2	1	2	1
10			Huber					3	1	2	1	3	1	2	1
11			Logistic Regression					3	1	2	1	3	1	2	1
12			Hinch and Talwar					4	1	4	1	4	1	4	1
13			Holland and Welsch					4	1	4	1	4	1	4	1
14		Regularisation	Lasso					2	0	2	0	2	0	2	0
15			Ridge Regression					2	0	2	0	2	0	2	0
16			Elastic Nets					2	0	2	0	2	0	2	0
17		Stepwise	Interaction	Bounded Steps				6	1	6	1	**7**	**1**	2	0
18				Unbounded Steps				4	1	6	1	1	0	2	0
19			Purequadratic	Bounded Steps				2	0	2	0	2	0	2	0
20				Unbounded Steps				6	1	5	1	5	1	2	1
21			Quadratic	Bounded Steps				6	1	6	1	6	1	6	1
22				Unbounded Steps				2	0	4	1	1	0	2	1
23			Polynomial	Bounded Steps				**13**	**1**	**13**	**2**	**7**	**1**	**13**	**2**
24				Unbounded Steps											

Table 6.18: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Double Supp. Predictors

Table 6.18: Score Table of the Average and Coefficient of Variation of the MAPE of Data Mining Models for Asynchronous Flow Lines including Elected Training Set (Sn) and Method (m) – Double Supp. Predictors (*cont.*)

								Train Set R_w at n=3,i=5		Train Set R_w at n=3,i=7		Train Set R_w at n=3,i=8	
								c		c^{-1}		$\log c$	
	Method				Model Specifications			Score		Score		Score	
Number (m)	Class I	Class II	Class III	Class IV	Data Mining Type	Learning Technique	Purpose	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$	$\mu Score_m$	$cScore_m$
1	Decision Tree	Boosting			Supervised Machine Learning	Classification	Comparison	1	1	1	1	1	1
2		Bootstrap Aggregating						1	0	1	0	1	0
3	Neural Network	Feedforward						1	0	1	0	2	0
4	Li and Meerkov (2009)				Current Formula (Literature)	N/A		13	1	13	1	13	1
5	$1/\mu_{\max}$							2	1	2	1	2	1
6	Multiple Linear Regression	Robust	Tukey's Bisquare		Supervised Machine Learning	Regression	Formula Derivation	1	1	1	1	5	1
7			Andrews					1	1	1	1	5	1
8			Cauchy M-estimators by Moore					1	1	1	1	5	1
9			Fair by Rey					1	1	1	1	2	1
10			Huber					1	1	1	1	3	1
11			Logistic Regression					1	1	1	1	3	1
12			Hinch and Talwar					1	1	1	1	5	1
13			Holland and Welsch					1	1	1	1	5	1
14		Regularisation	Lasso					0	0	0	0	2	0
15			Ridge Regression					2	0	2	0	2	0
16			Elastic Nets					0	0	0	0	2	0
17		Stepwise	Interaction	Bounded Steps				6	1	6	1	7	1
18				Unbounded Steps				0	0			2	0
19			Purequadratic	Bounded Steps				2	0	2	0	2	0
20				Unbounded Steps				1	1			6	1
21			Quadratic	Bounded Steps				6	1	6	1	7	1
22				Unbounded Steps				1	1	1	1	2	0
23			Polynomial	Bounded Steps				**13**	**2**	**13**	**2**	**13**	**1**
24				Unbounded Steps									

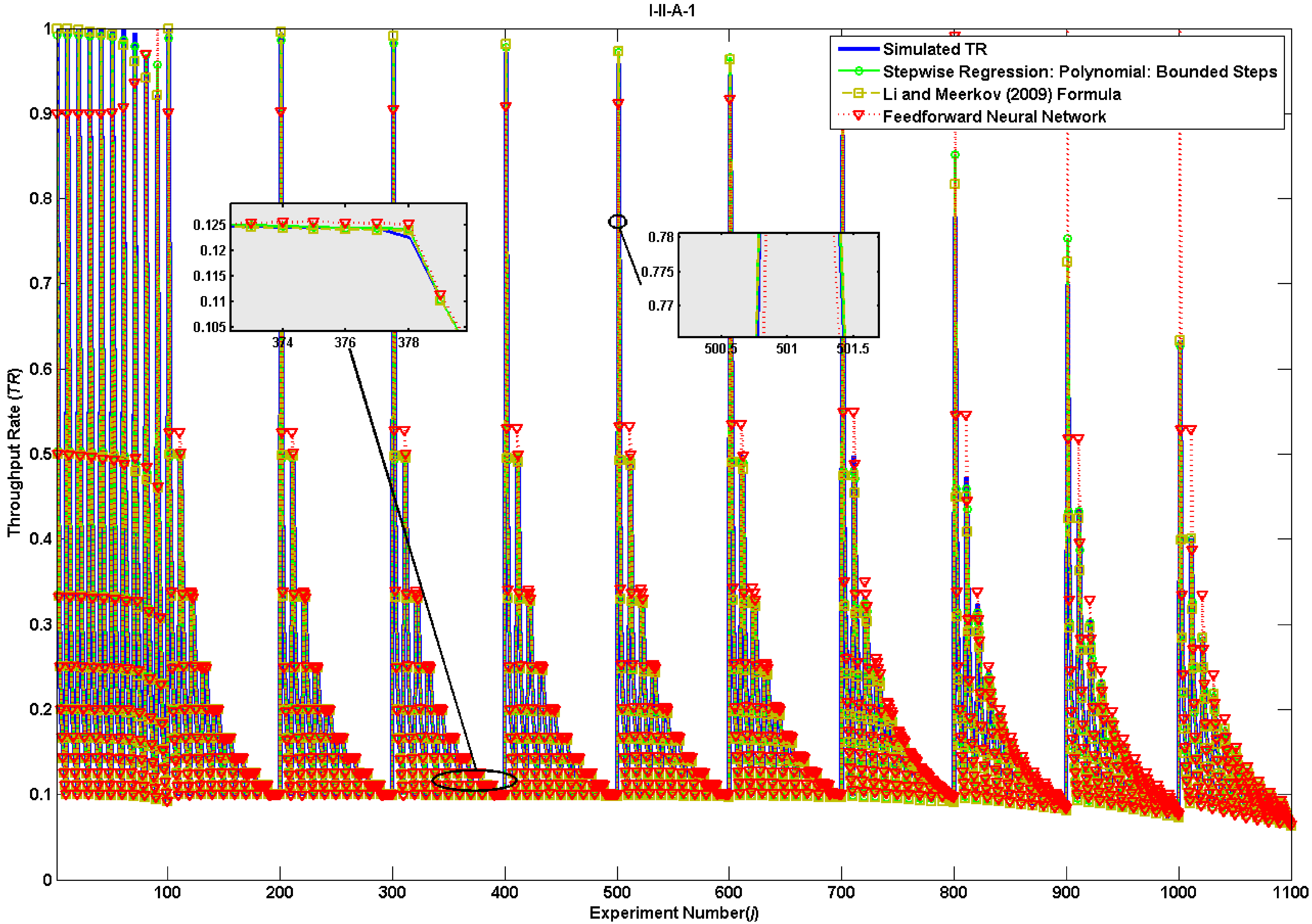


Figure 6.13: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-A-1$

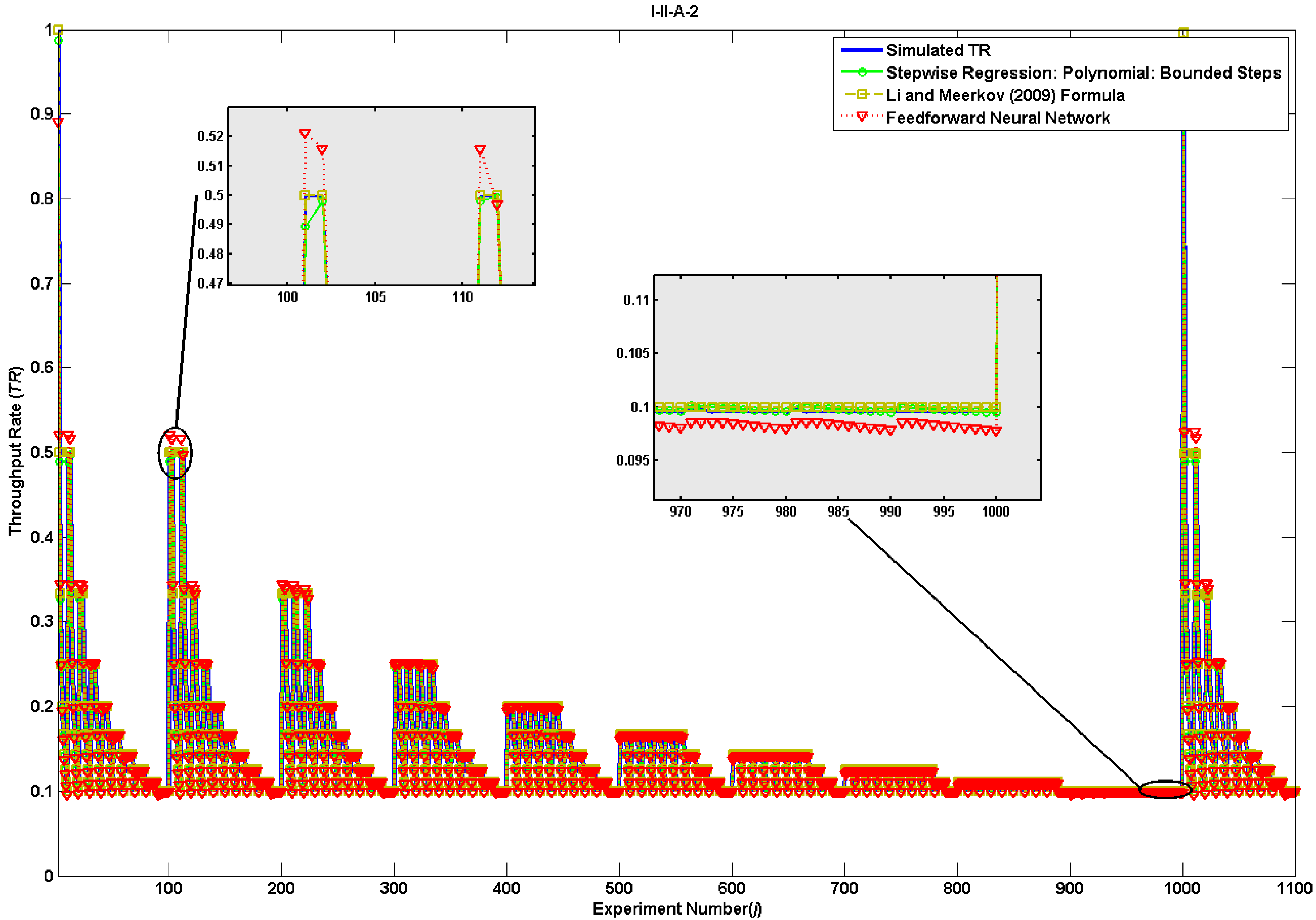


Figure 6.14: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-A-2$

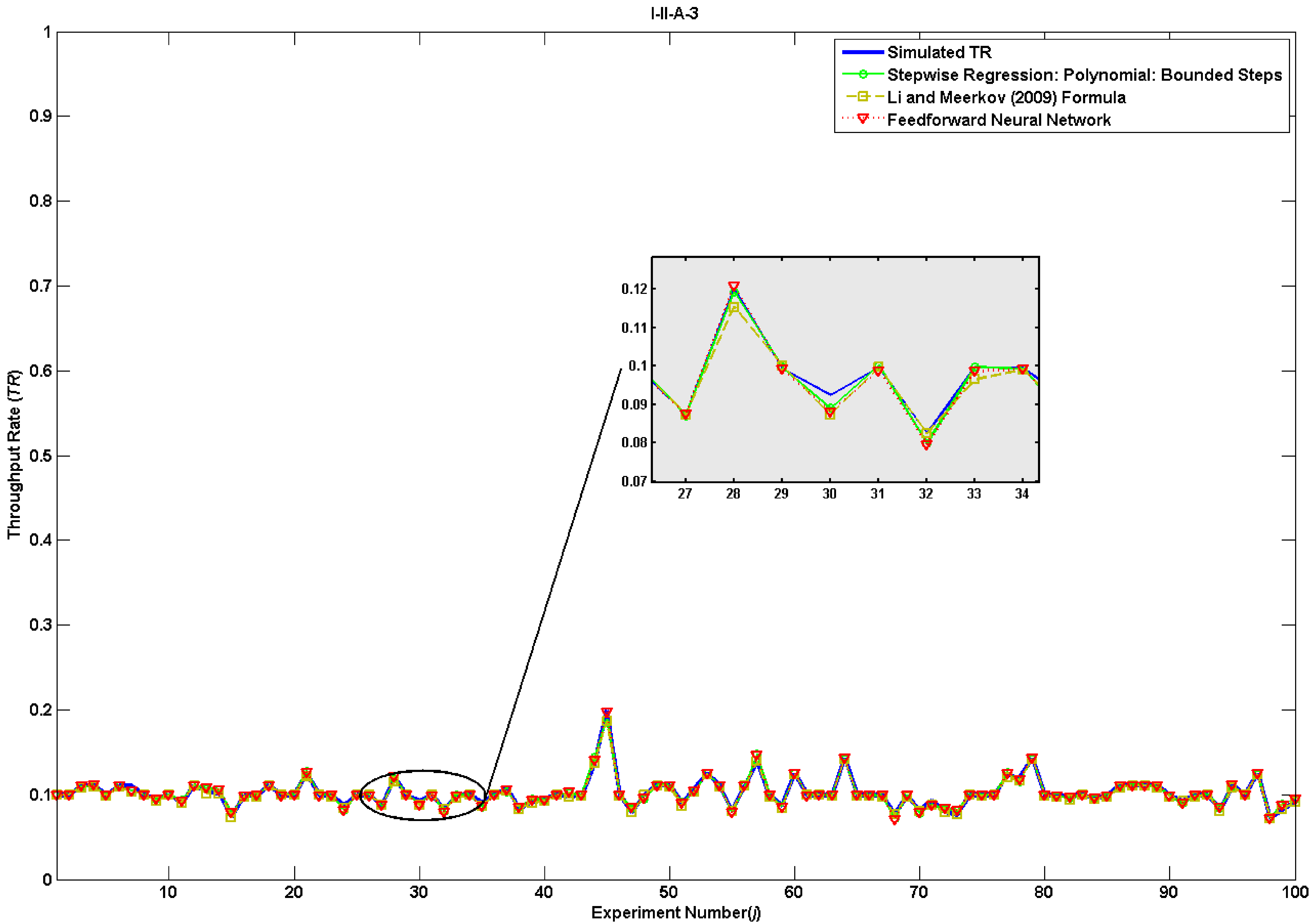


Figure 6.15: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-A-3$

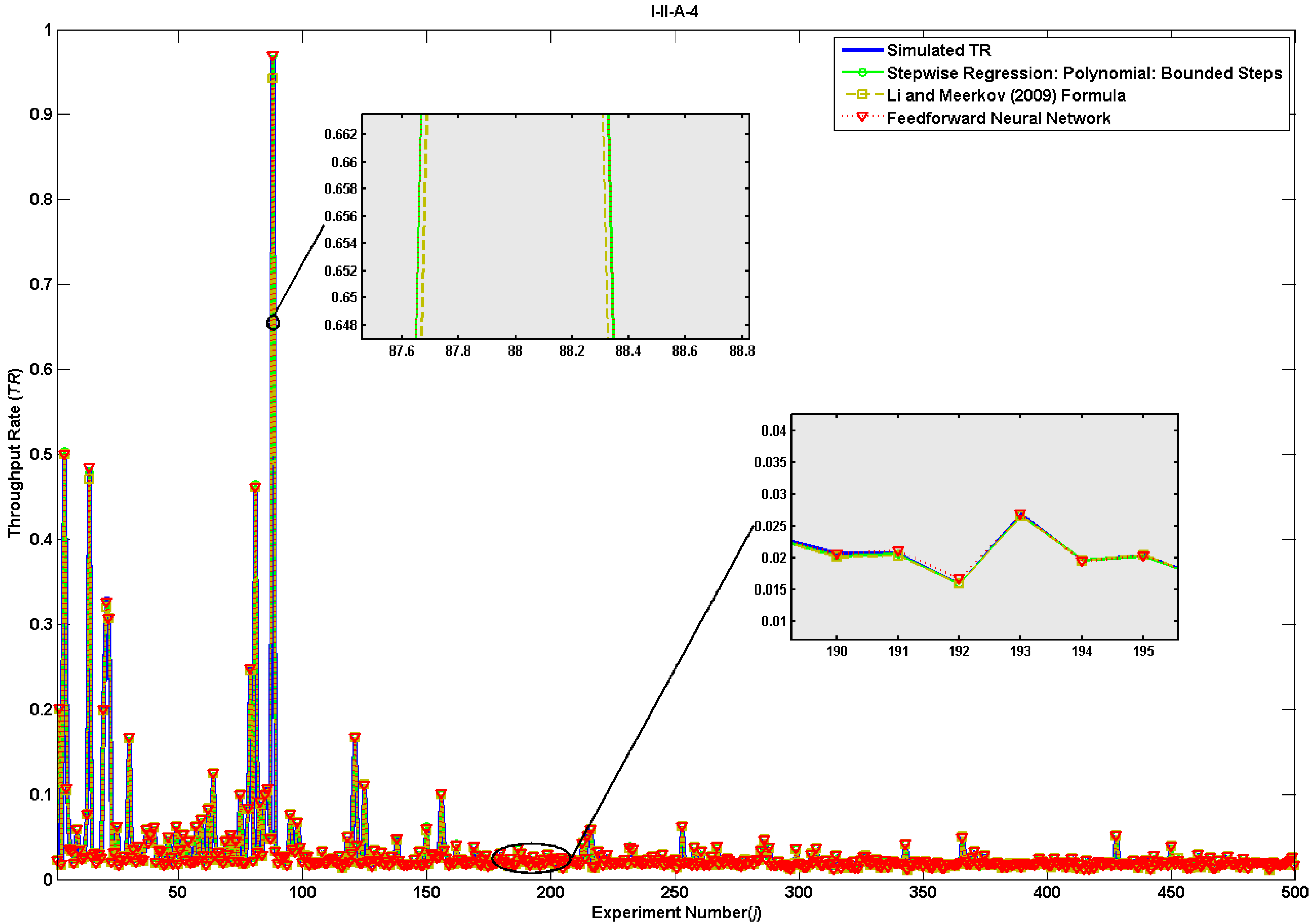


Figure 6.16: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-A-4$

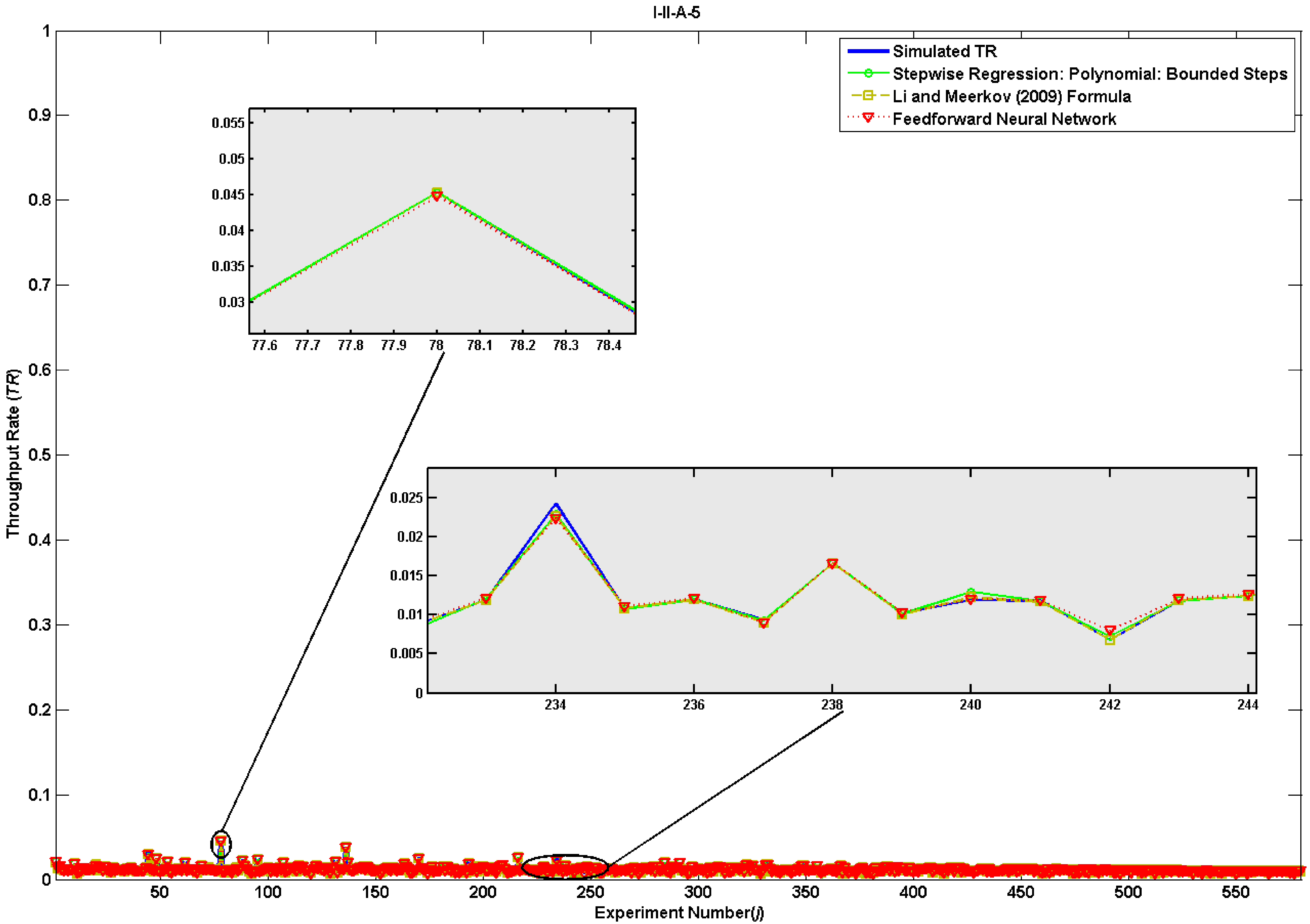


Figure 6.17: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-A-5$

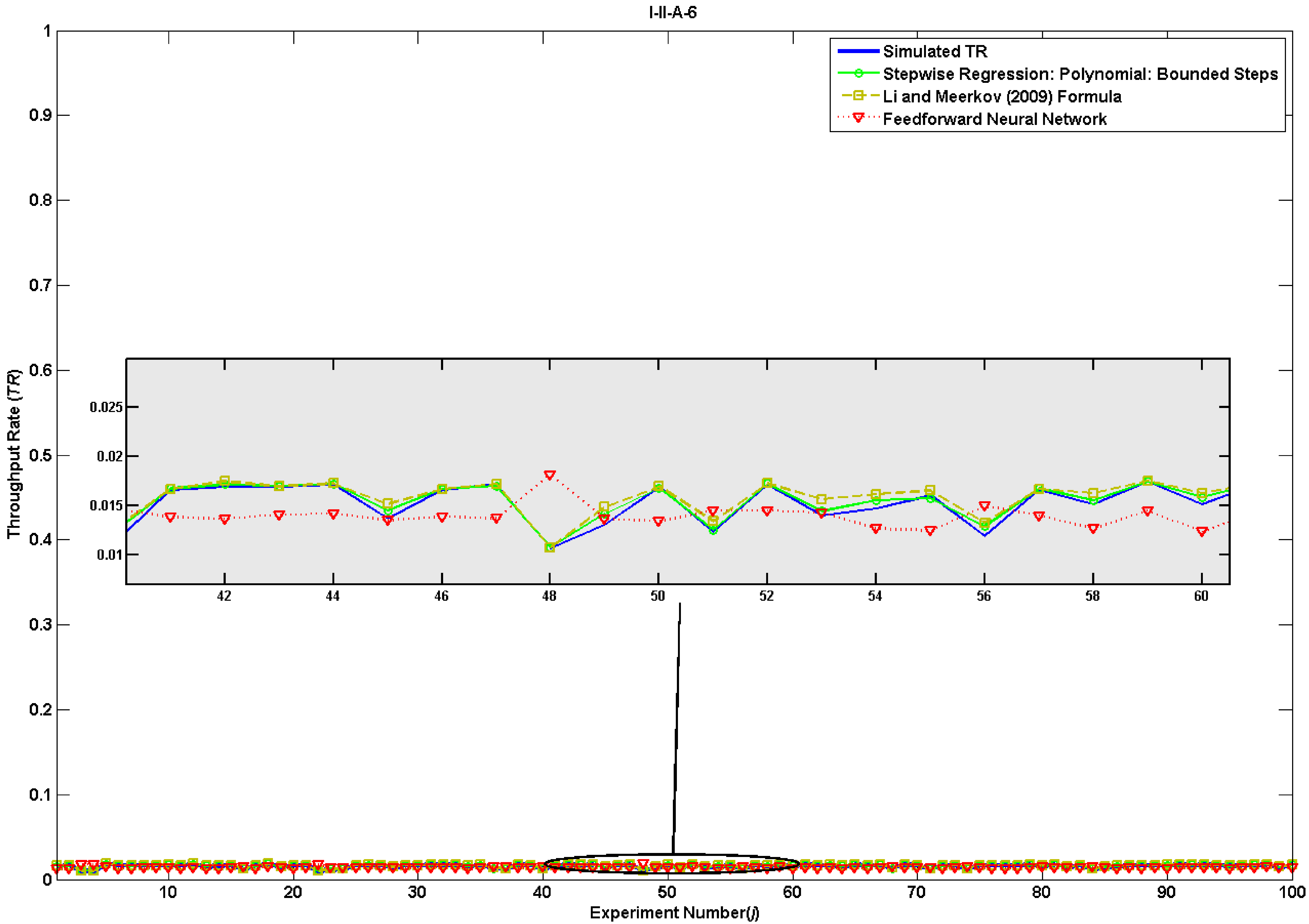


Figure 6.18: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-A-6$

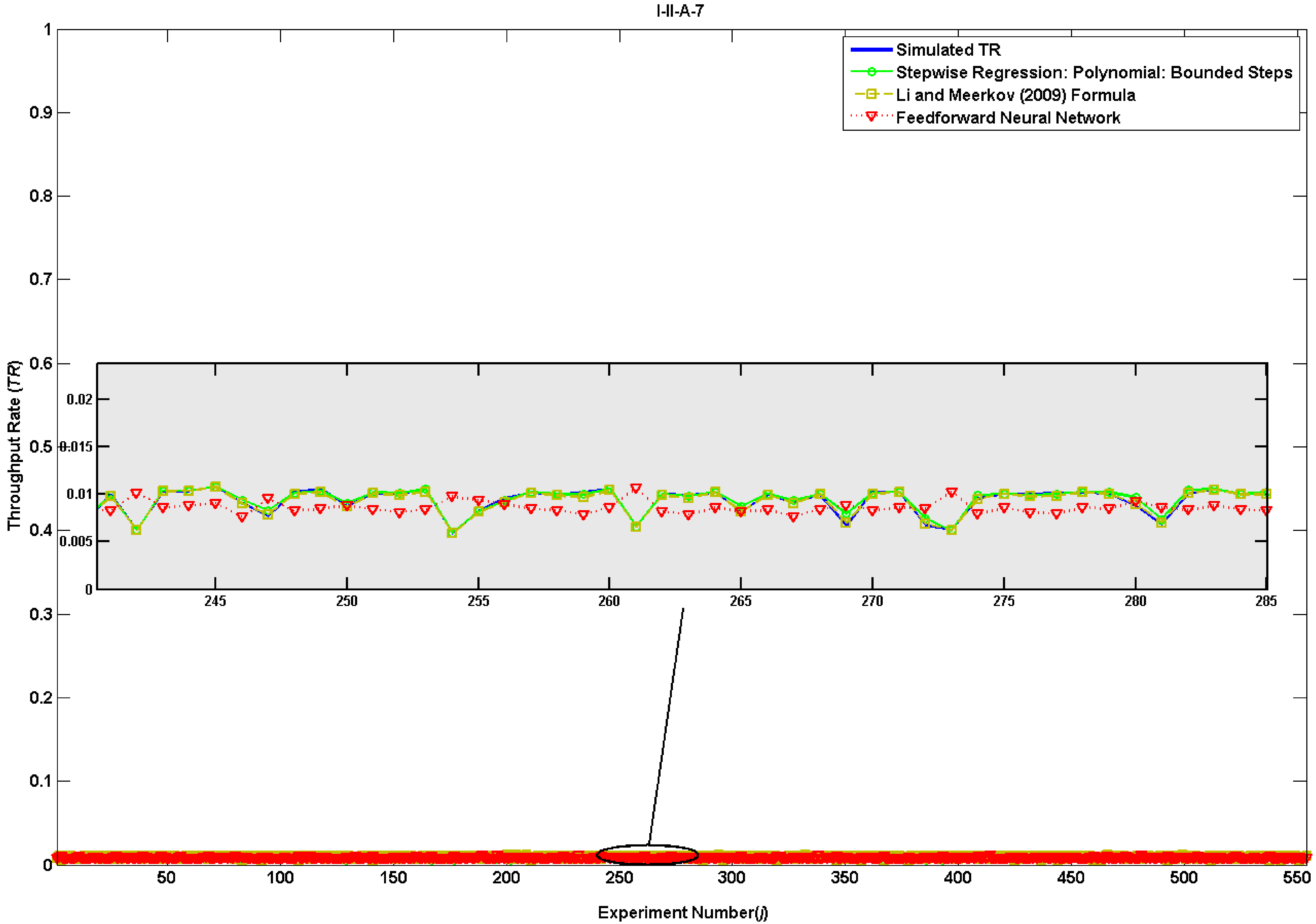


Figure 6.19: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-A-7$

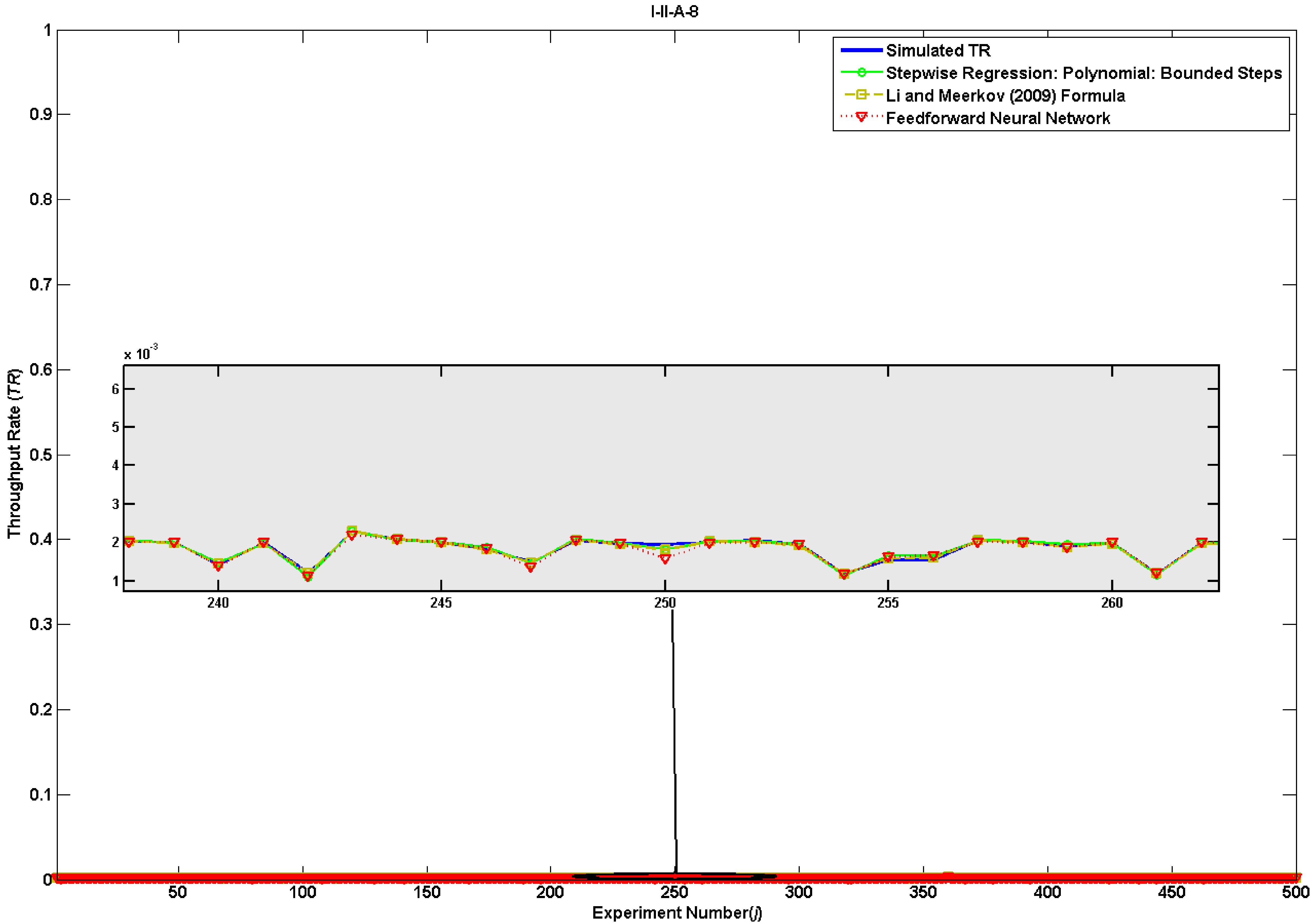


Figure 6.20: Predicted Throughput Rate of the Optimal Regression Model against Simulated Throughput Rate and Comparison Models for Test Data Set $I-II-A-8$

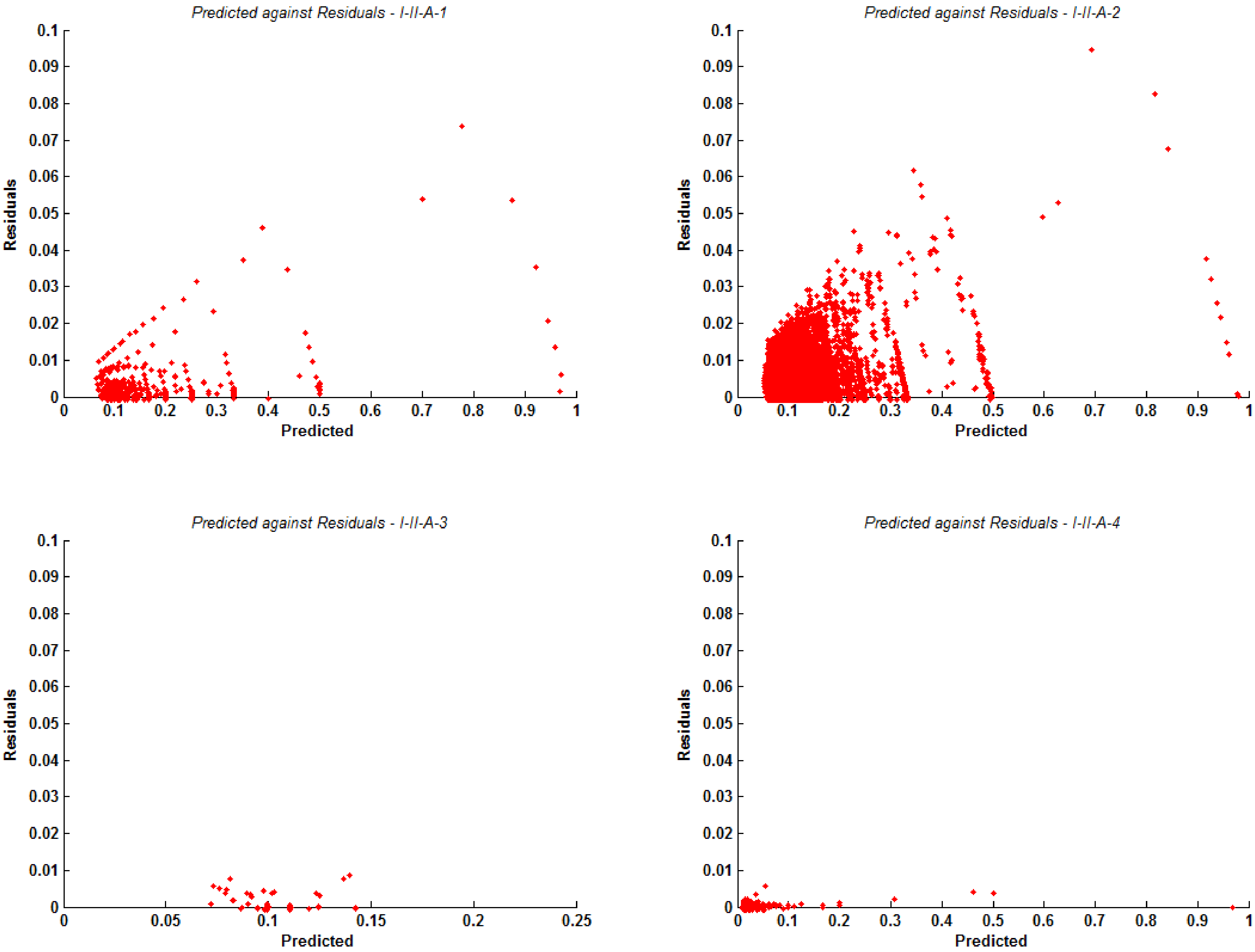


Figure 6.21: Predicted against Residuals Plots using the Optimal Regression Model for Test Data Sets $I-II-A-1$ **to** $I-II-A-8$

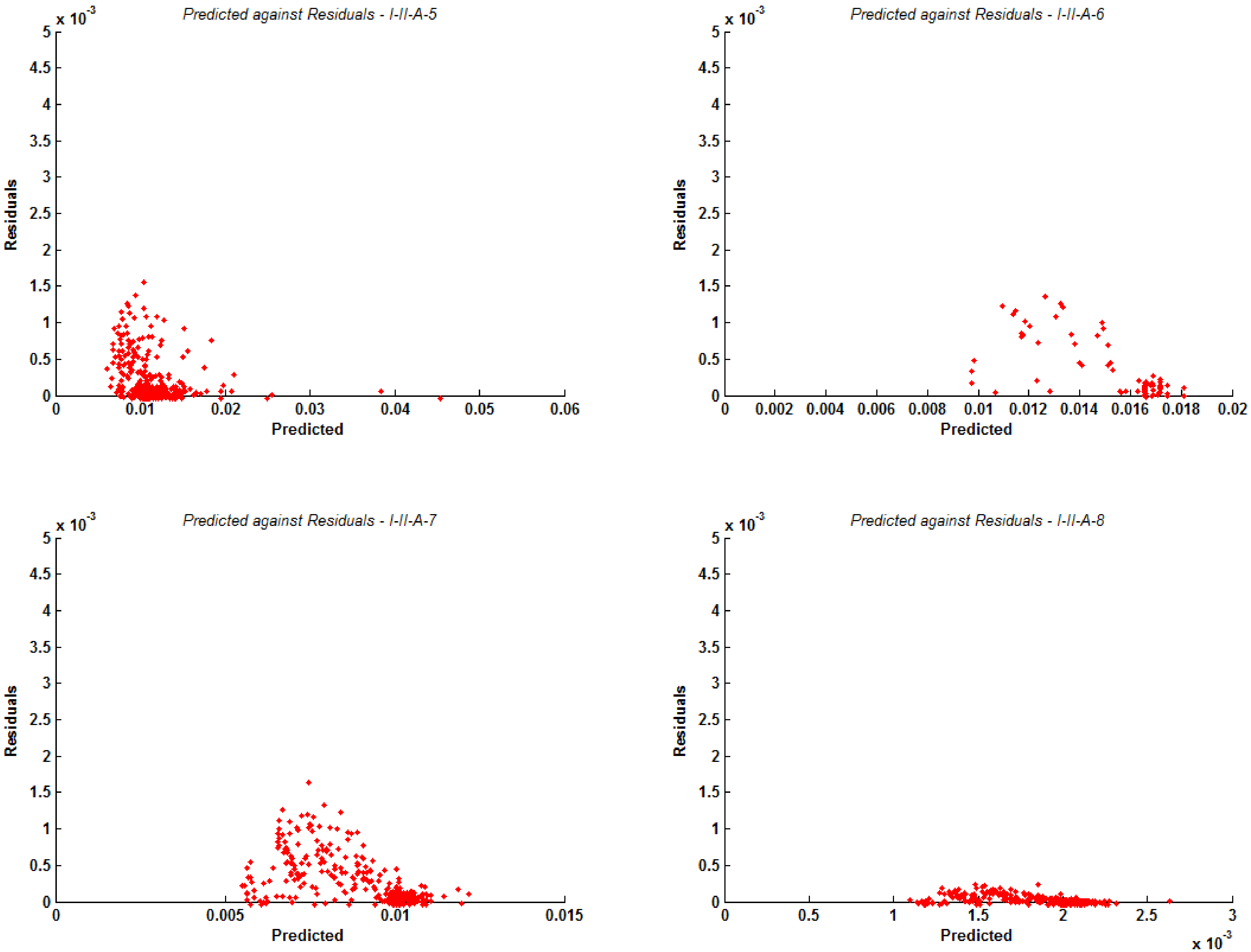


Figure 6.21: Predicted against Residuals Plots using the Optimal Regression Model for Test Data Sets $I-II-A-1$ to $I-II-A-8$ (*cont.*)

6.5 Empirical Formula Validation

The empirical formula was applied to the real-world case study to determine the calculated throughput rate for variability scenarios. The stochastic nature of this construction project is presented in terms of 6 variability factors. The controllability of these factors changes based on their nature and the process owner. Factors related to processes owned by Costain can be fully controlled while for the processes owned by their contractors require co-ordination between process owners, hence, semi controllable. Factors outside the control of process owners, e.g., the traffic congestions, are referred to as uncontrollable. In terms of optimisation, the first two categories of variability factors, i.e., controllable and semi-controllable factors, fall under the '*decision variables'* while the third is constraints. Table 6.17 lists the variability factors and the conditions and range of each factor.

Table 6.19: Decision Variables and Constraints for the Real-world Case Study

Decision Variables

Factor		Lower Bound	Upper Bound	Type	Step
Arrival Rate (1/min)	Mean	5	80	Discrete	1
	Standard Deviation	0	80	Discrete	0.1
Load Size (m3)		6	8	Discrete	0.5
Concrete Plant		1	3*	Discrete	1
No. of Deliveries from the Same Concrete Plant		1	2	Discrete	1
Mean Site Delay (min)		0	10	Discrete	1

Constraints

Process/ Queue	Description
Batch and Load	i. Max. two trucks can be batched at the same time, i.e., two lanes; and ii. Max. queue time at concrete plant is 10 min.
Drive to Site	Delivery time changes throughput the day based on traffic congestions.
Site Access	The site access allows only one truck to pass through at a time.
Add Water/ High Slump Load Queue	Trucks route out to these processes/queue based on pre-determined pattern based on historical data.
Discharge and Extrude	A 30 minutes break for refuelling at a specific time is enforced; during that time the extruder machine will complete any hold trucks before block routing in until refuelling is complete.

**3 indicates that both concrete plants are used*

Subsequently, the number of decision variables was increased gradually to cover all possible combination of decisions that need to be taken at a particular day (Table 6.18). In terms of pre-processing of the decision variables, the arrival rate and site delay factors are straightforward as they are presented by the processing time. The load size, however, affects the batch and discharge processing times as outlined in Section 5.3.2.2. Likewise, concrete plant decides on the delivery time. Finally, the number of deliveries from the same concrete plant determines if a parallel processing of the '*Drive to Site*' process is allowed. The predictor terms $\mu_{\max}{}^{-1}$, $e^{c_{av}}$, N^{-1} and $e^{\mu^{-1}}$ were then obtained for each variability scenario to feed the formula and the throughput rate was calculated. Figure 6.22 shows the calculated against the simulated throughput rate for the variability scenarios of one of the decision combination scenarios, i.e., Number 6. The results show that the calculated TR follows closely the simulated TR. However, there are few oscillations in the calculated TR which are not present in the simulated TR. The presence of such oscillations can be due to the introduction of continuous data in the validation case study which were not present in the test sets $(I/II-A-1$ to $I/II-A-8)$. The prediction accuracy, presented by MAPE, MINAPE and MAXAPE, for the same scenario is shown in Table 6.19. The correlation analysis between variability of the main activities and the residuals is presented in Table 6.20.

Table 6.20: Scenarios of all Possible Decision Combinations

Scenario No.	Arrival Rate	Load Size	Concrete Plant	No. of Deliver.	Site Delay
1	√	X	X	X	X
2	√	√	X	X	X
3	√	X	√	X	X
4	√	X	X	√	X
5	√	X	X	X	√
6	√	√	√	X	X
7	√	√	X		X
8	√	√	X	X	√
9	√	X	√	√	X
10	√	X	√	X	√
11	√	X	X	√	√
12	√	√	√	√	X
13	√	√	√	X	√
14	√	√	X	√	√
15	√	X	√	√	√
16	√	√	√	√	√

√ *Decision Variable*

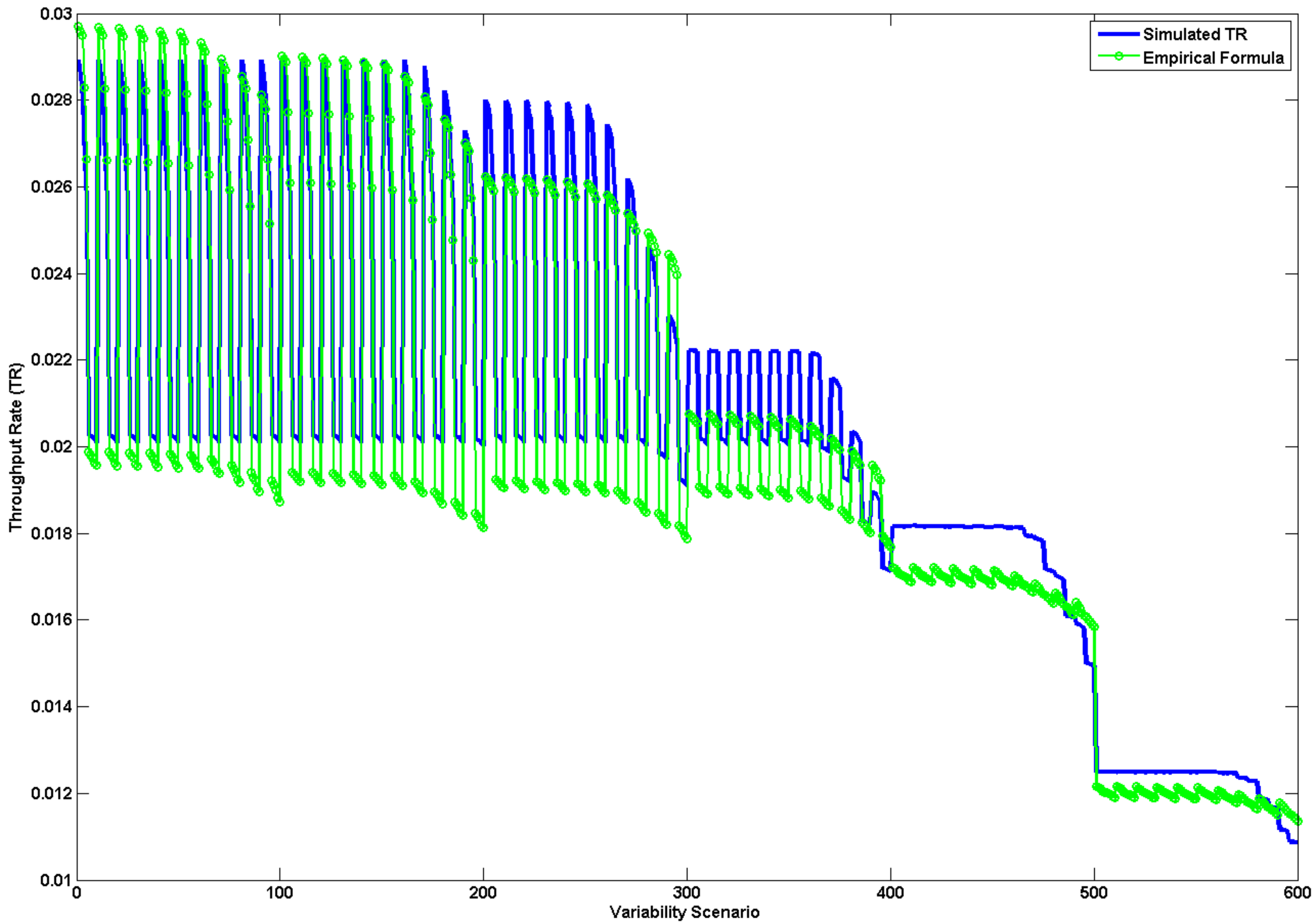


Figure 6.22: Predicted Throughput Rate using Formula against Simulated Throughput Rate for Decision Combination Scenario No. 6 of the Real-world Case Study

The slowest process, i.e., truck delivery, followed by the arrival rates, which are highly variable, are the closest to explain the prediction error of the empirical formula, however, the correlation coefficients for them are still low.

The correlation between the range of variation in mean processing time and coefficient of variation and the throughput rate was also tested to investigate if the formula will remain competent for other continuous data. This was done using the regression formula covariates in Equation 6.4, expect N^{-1} since the length is constant, as they represent the mean processing time and coefficient of variation terms with highest relationship to the throughput rate. The correlation analysis (Table 6.21) showed that the change in MAPE of the throughput rate is not highly related to changes in the mean processing time and coefficient of variation terms. This suggests that the formula performance remains valid to other case studies with different continuous data, however, a slight decay in performance might occur since all correlation coefficients are negative. Figure 6.23 shows the change in prediction error across the range of $\mu_{\max}, c_{av}$ and μ.

Table 6.21: Prediction Accuracy of the Empirical Formula

	Empirical Formula
Mean Absolute Error Percentage (MAPE)	4.470%
Minimum Absolute Error Percentage (MINAPE)	0.035%
Maximum Absolute Error Percentage (MAXAPE)	9.449%

Table 6.22: Correlation Coefficient between Residuals and the Main Activities

	Arrival Station	*Batch Station*	*Truck Delivery*	*Load Condition*	*Discharge*
Residuals	0.16	0.06	0.25	0	0.06

Table 6.23: Correlation Coefficient between Residuals and the Mean Processing Time and Coefficient of Variation Terms

	$\mu_{\max}^{-1}$	$e^{c_{av}}$	$e^{\mu^{-1}}$
Residuals	-0.25	-0.15	-0.31

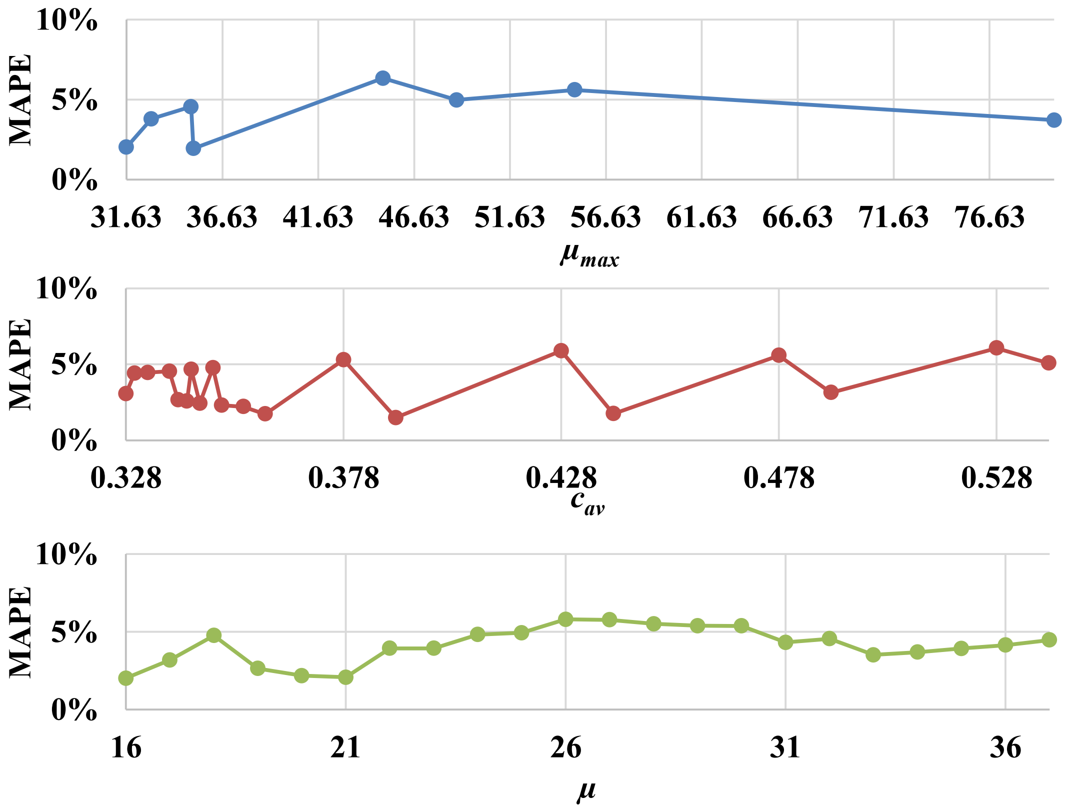


Figure 6.23: MAPE for the Range of $\mu_{\max}, c_{av}$ and μ of the Variability Scenarios of the Real-world Case Study

6.6 Summary

In conclusion, statistical analysis showed that for asynchronous flow lines, $\mu_{\max}{}^{-1}$, c_{av}, $e^{c_{av}}$, N and N^{-1} are significant and $e^{\mu_{\min}{}^{-1}}$, μ, μ^{-1}, $\log\mu$, $e^{\mu^{-1}}$, c, c^{-1} and $\log c$ are potentially significant, hence, classified as main and free predictions, respectively, for model building. These reduced to the equivalent parameter terms, i.e., c, e^{c}, μ^{-1}, N and N^{-1}, for synchronous flow lines.

For model building of synchronous flow lines, one full factorial DOE-based training data set was sufficient to reach to an optimal regression model. Whereas, asynchronous flow lines required three data sets, each includes the main predictors and $e^{\mu^{-1}}$.

Polynomial regression with bounded steps was the best performer among all classification and regression models and it provided both less prediction errors and stability across the range of test data sets. Stepwise regression outperformed the robust and regularisation algorithms, which suggests that not only the ability to add and remove predictor terms is important, as in regularisation algorithms, but also multiplication and higher order terms can improve the model performance. Furthermore, forward iteration of stepwise regression usually leads to better prediction accuracy and smaller model size than backward iteration, if both have the same freedom in covariate terms.

Developed regression models for the throughput rate of synchronous and asynchronous flow lines surpass the performance of the best comparison models, i.e., Blumenfeld and Li and Meerkov formula, respectively. For the asynchronous flow lines, MAPE increased from 2% with discrete data to 5% with continuous data from the real-world case-study. Next chapter investigates the use of this regression model to build an autonomous control method.

7 AUTONOMOUS CONTROL FRAMEWORK

7.1 Introduction

This chapter presents the results related to the second methodological framework of this research, i.e., Autonomous Control Framework. The empirical formula (Equation 6.4) formulates the relationship between the asynchronous flow line variability, represented by line-based predictor terms derived from the normal distribution parameters, and the throughput rate. The formula was used within the Autonomous Control Framework, according to the logic explained in Section 4.5.3.3, to examine the future state of the system under different variability scenarios and autonomously control the decision variables of the system to optimise the throughput rate.

Two asynchronous flow lines were investigated. The first is a 3x3 flexible flow line model developed by Scholz-Reiter (2005) with representative examples of variability. The second is the real-world case study described in Section 5. Throughout these case studies, the model was compared with existing autonomous and optimisation techniques for validation reasons.

7.2 3x3 Flexible Flow Line

The Queue Length Estimator (QLE) method was applied to the 3x3 flexible flow line described in Scholz-Reiter et al. (2005) with seasonal fluctuation of orders to reproduce the results and ensure that the model is accurately represented and the method is applied correctly (Figure 7.1). As shown in Table 7.1, apart from the standard deviation of the throughput time, the original and reproduced results match to a good accuracy.

The formula-based autonomous control method was then compared against the QLE and Past Events Based (PEB) autonomous control and OptQuest optimisation techniques. The same 3x3 flow line model was used but with different representative scenarios of variability as shown in Table 7.2.

The first set '*Stochastic Process*' consists of several scenarios of processing times variability. In synchronous flow lines (sync), the mean processing times are kept constant across all processing stages for each product type. The coefficient of variation is the same for all product types across the whole flow line. As for asynchronous lines (async), the mean processing time μ, coefficient of variation c or both change from one processing stage to another for each individual product.

In this case c is process dependent, apart from; scenario nine, where c is constant for all product types and scenario ten with independent c. The system was overloaded with arrival rates of the input source significantly lower than the processing times to test how the autonomous methods will handle such a dynamic situation. Through all scenarios, each individual process is optimised, in terms of mean processing times, to a single product.

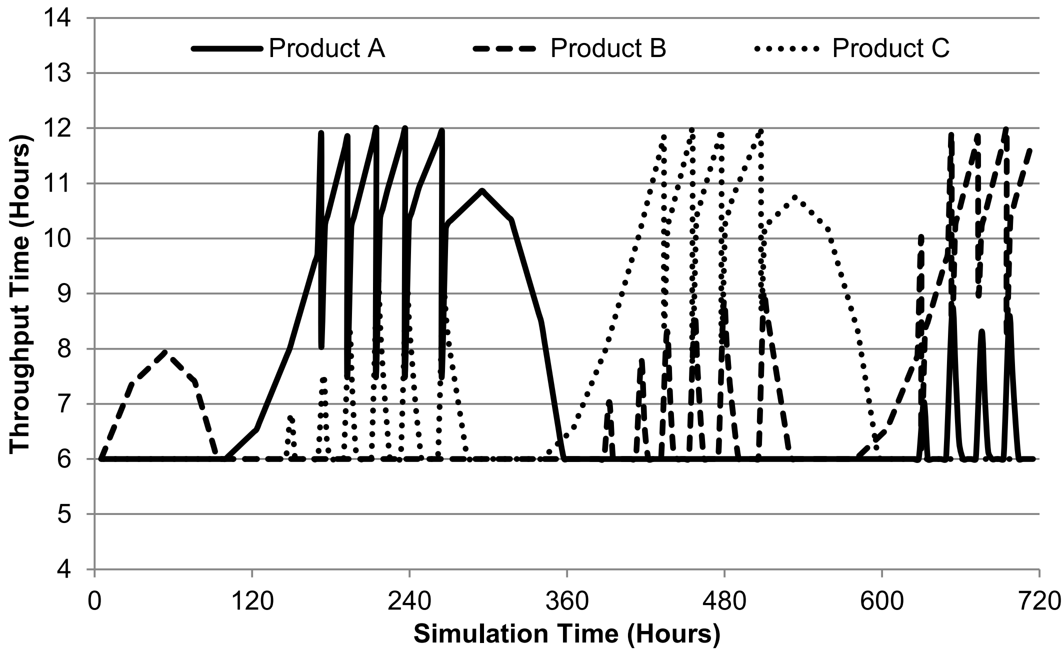


Figure 7.1: Reproduction of Throughput Time with QLE Autonomous Control on Scholz-Reiter et al. (2005) model

Table 7.1: Comparison of the Original and Reproduced Throughput Time with QLE Autonomous Control

Parameter	Original Results	Reproduced Results	Deviation	
Min Throughput Time (Hours)	6	6	0	0%
Mean Throughput Time (Hours)	6.77	7.51	+0.74	+11%
Max Throughput Time (Hours)	12.28	11.99	-0.29	-2%
Standard Deviation of Throughput Time (Hours)	1.12	1.99	+0.87	+78%

Table 7.2: Representative Scenarios used for Performance Analysis

Stochastic Process

No.	**Flow line Type**	**Processing Time (unit)**		**Arrival Rate (1/unit)**	
		μ	c	μ_a	c_a
S-1	Synchronous (Intra-process variability)	6-10 (sync)	0 (sync)	1	0
S-2		6-10 (sync)	0.025 (sync)	1	0
S-3		6-10 (sync)	0.5 (sync)	1	0
S-4		6-10 (sync)	1 (sync)	1	0
S-5	Asynchronous (Intra- and inter-process variability)	2-10 (async)	0.5 (sync)	1	0
S-6		6-10 (sync)	0.4-0.8 (async)	1	0
S-7		2-10 (async)	0.4-0.8 (async)	1	0
S-8		2-10 (async)	0.3-0.9 (async)	1	0
S-9		2-10 (async)	0.1-0.8 (async)	1	0
S-10		2-10 (async)	0.02-1 (async)	1	0

Arrival Rate

No.	**Flow line Type**	**Processing Time (unit)**		**Arrival Rate (1/unit)**	
		μ	c	μ_a	c_a
S-7/A-1	Asynchronous (Intra- and inter-process variability)	2-10 (async)	0.4-0.8 (async)	1	0
A-2		2-10 (async)	0.4-0.8 (async)	5	0
A-3		2-10 (async)	0.4-0.8 (async)	50	0
A-4		2-10 (async)	0.4-0.8 (async)	1	0.5
A-5		2-10 (async)	0.4-0.8 (async)	5	0.5
A-6		2-10 (async)	0.4-0.8 (async)	50	1

The arrival rates in the second set of Table 7.2 were investigated but as normally distributed arrival rates to match the developed empirical formula. The mean arrival rates μ_a were set extremely low, 80% of average of μ to replicate actual static flow lines as suggested by Scholz-Reiter et al. (2005), and significantly high in comparison to μ for dynamic situation covering both underloaded and overloaded systems, with coefficient of variations c_a changing in the same manner.

Windt and Becker (2009) identified four performance measures for autonomous control (Section 2.5). These performance measures were analysed, apart from the Due Date which is outside the research scope. As for the simulation-based optimisation using OptQuest, maximising the average throughput was set the optimisation objective.

Figure 7.2-7 shows a comparison of autonomous control and optimisation techniques based on the DES modelling results.

The left column '*Stochastic Process*' scenarios show the effect of processing times variability of individual process on the performance measures. As expected, the queue and throughput times have direct relationship with the change in *c*. Utilisation on the other hand is inversely related to *c* but the change is not significant. Reduction of processing times decreases the accumulated queues and throughput time regardless of c even in the case of asynchronous flow line.

The developed technique performed well on all the synchronous and asynchronous flow line scenarios, followed by QLE autonomous control logic for synchronous flow lines. Autonomous control based on past events (PEB) performed poorly for these scenarios and a simple circulation of products (Nil) gave better results. PEB performance improved to reach almost the same levels as QLE, for both mean queue and throughput queue times, when the inter-variability of the mean processing time is introduced into the system. It is worth mentioning that the same does not apply with introduction of inter-variability of the coefficient of variation solely ('*Stochastic Process*' Scenario S-6). Performance curves for standard deviation of queue and throughput times follow closely the ones of mean apart from the PEB autonomous control; its performance still directly related to inter-variability of mean processing time but it did not reach the same levels as QLE.

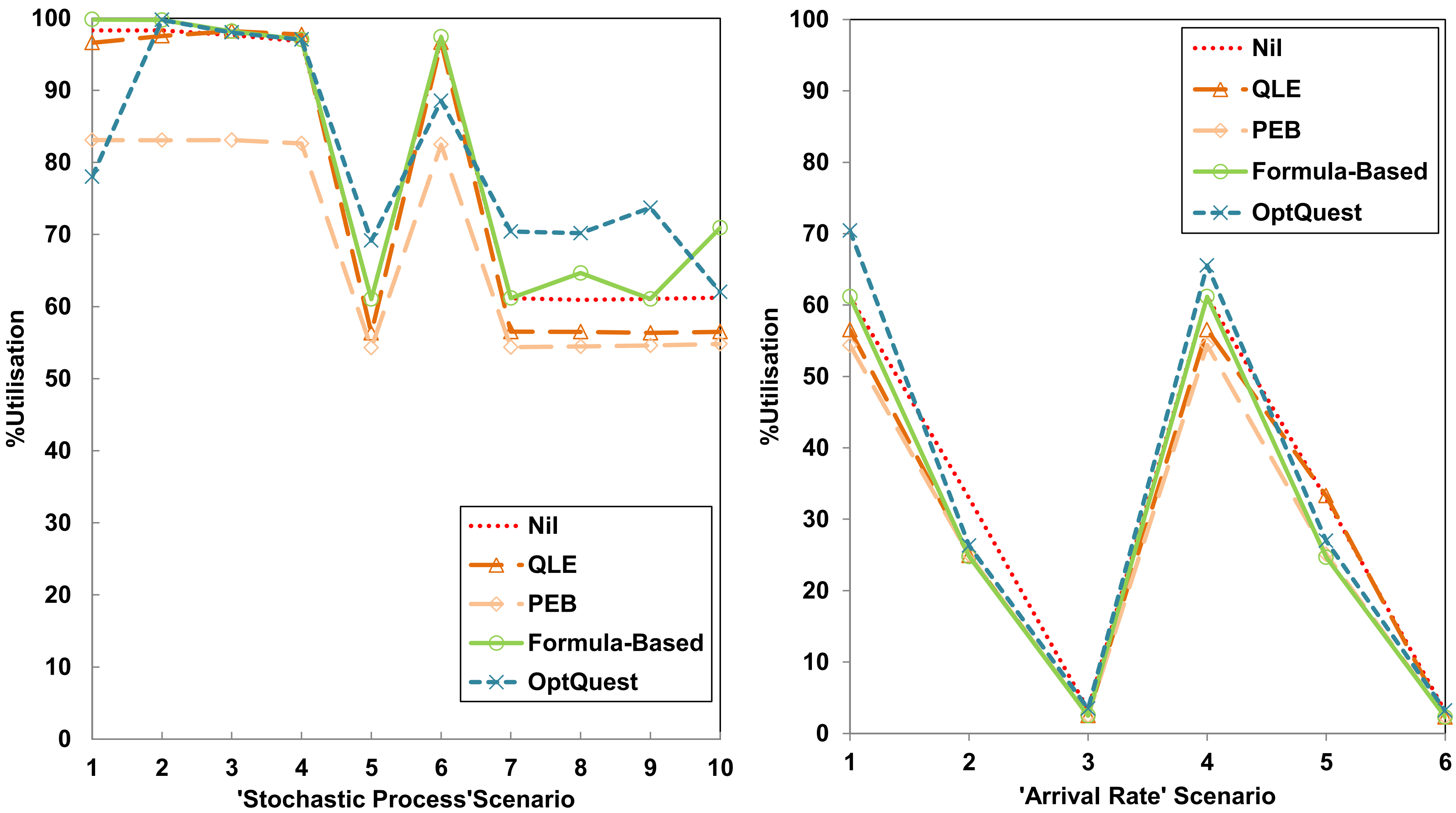


Figure 7.2: Performance Analysis on Representative Scenarios - Utilisation

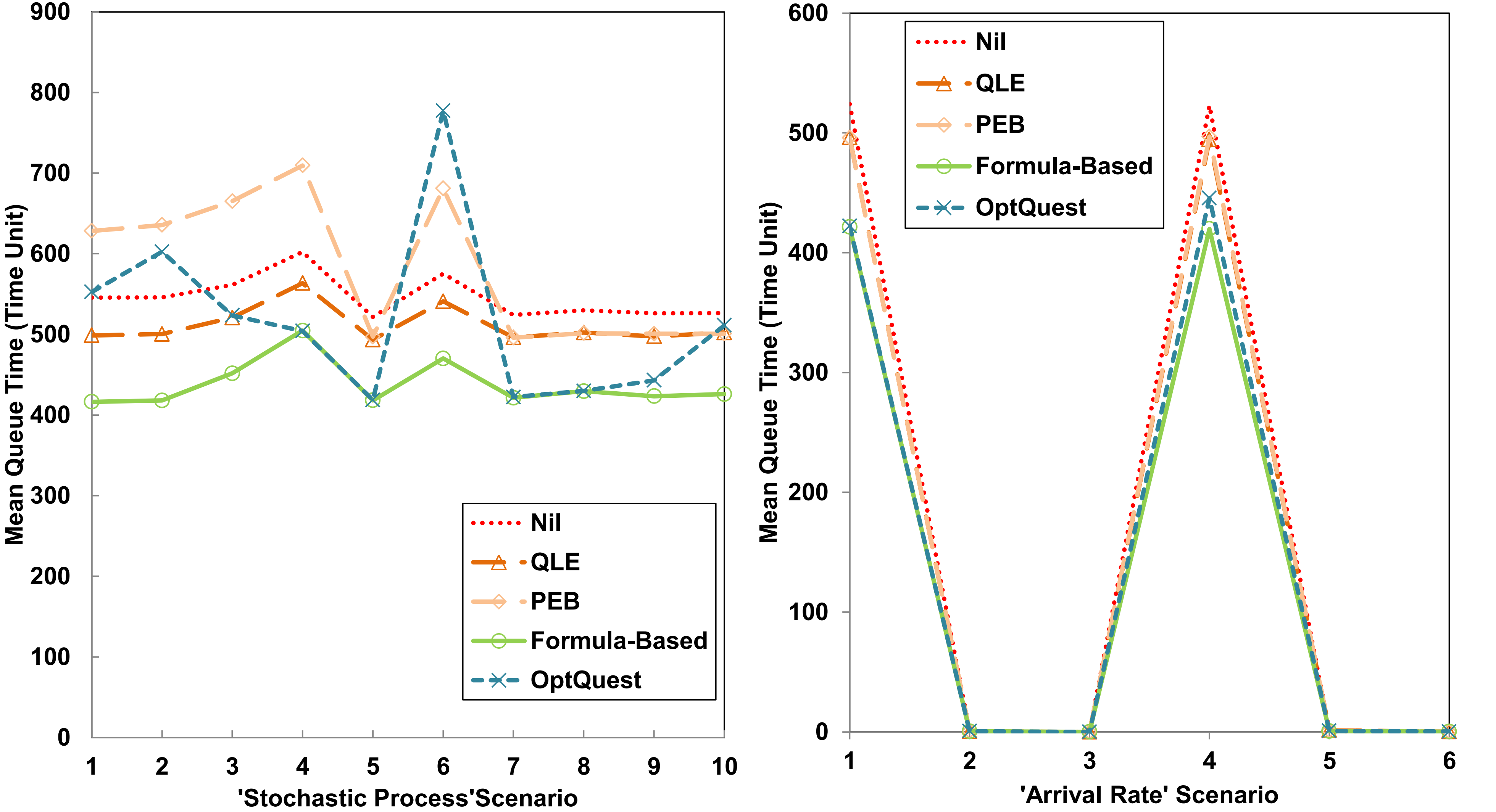


Figure 7.3: Performance Analysis on Representative Scenarios – Mean Queue Time

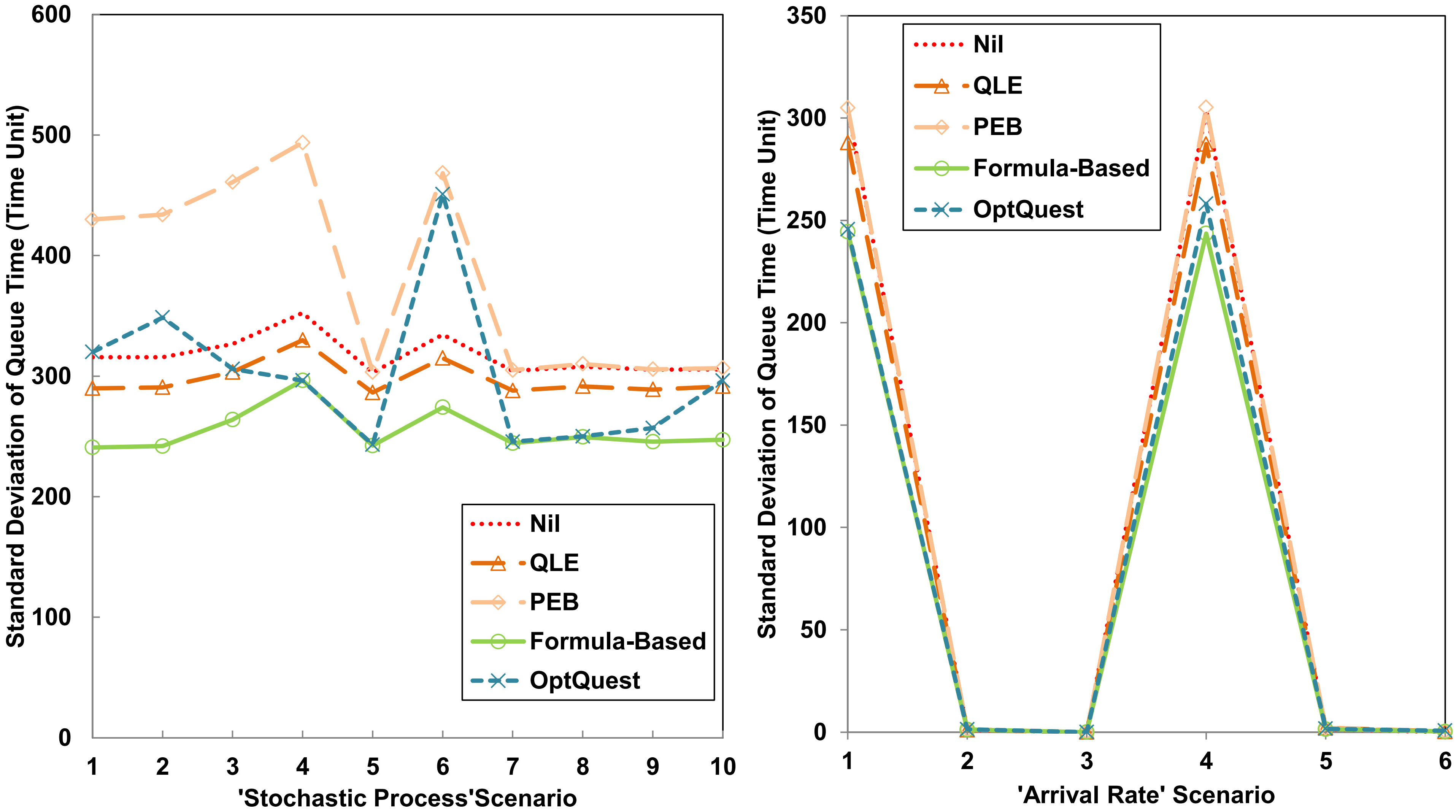


Figure 7.4: Performance Analysis on Representative Scenarios – Standard Deviation of Queue Time

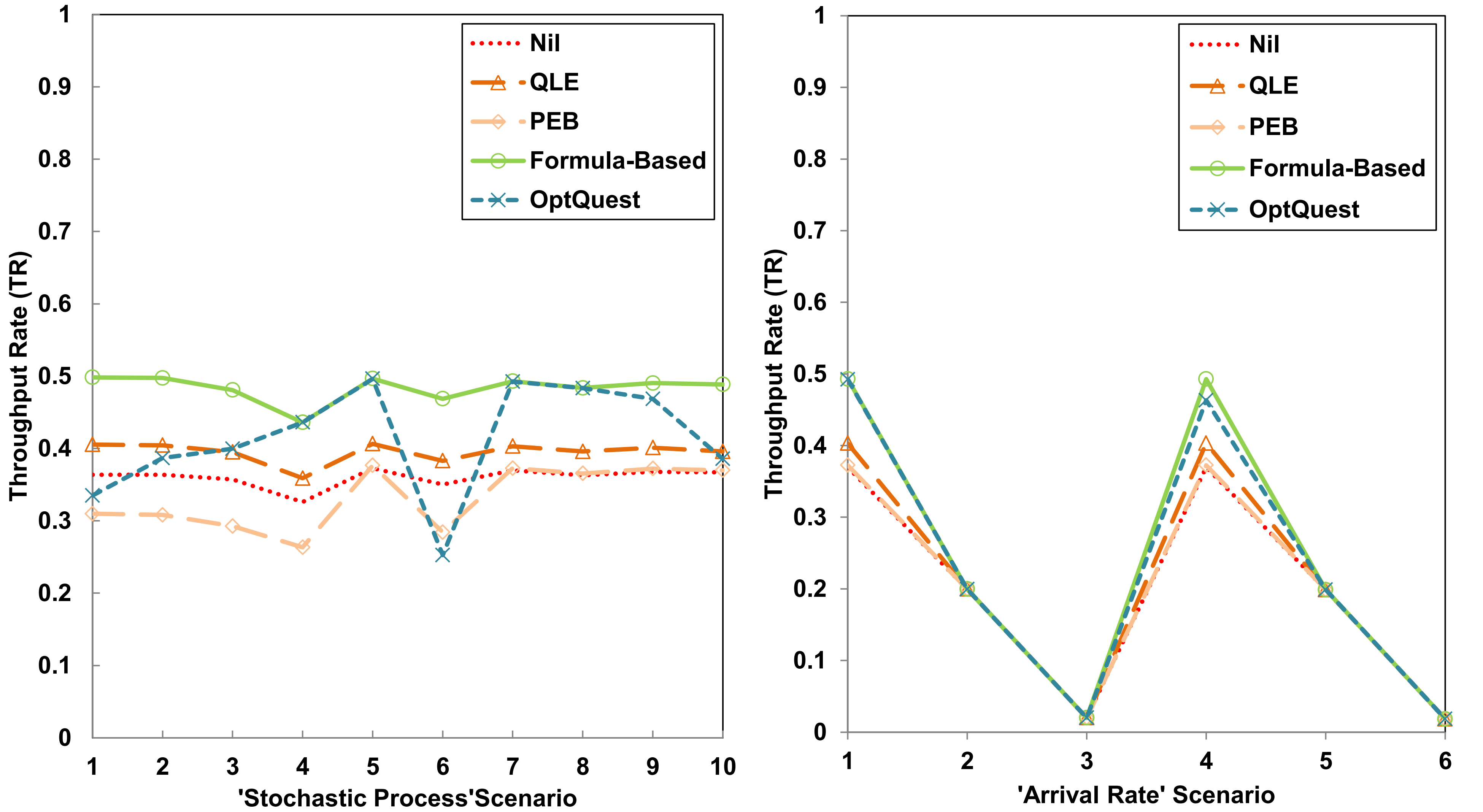
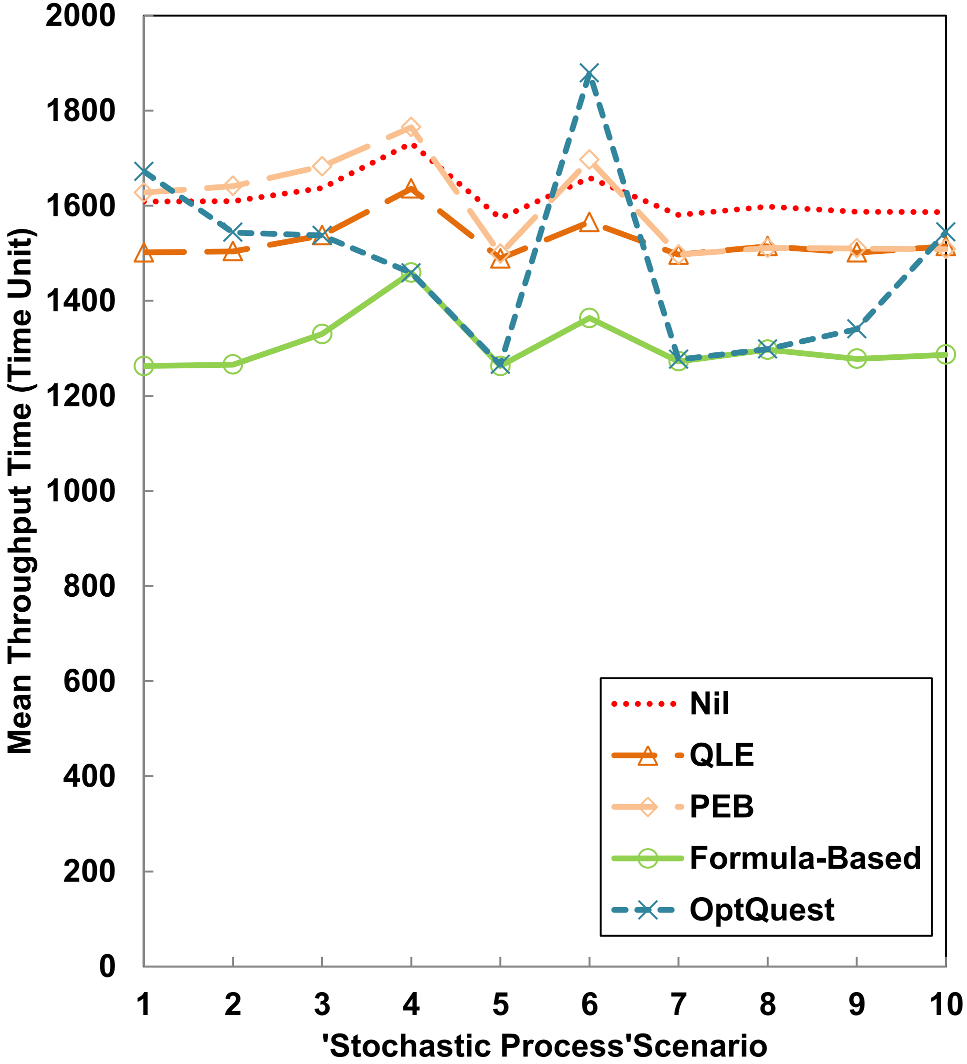


Figure 7.5: Performance Analysis on Representative Scenarios – Throughput Rate



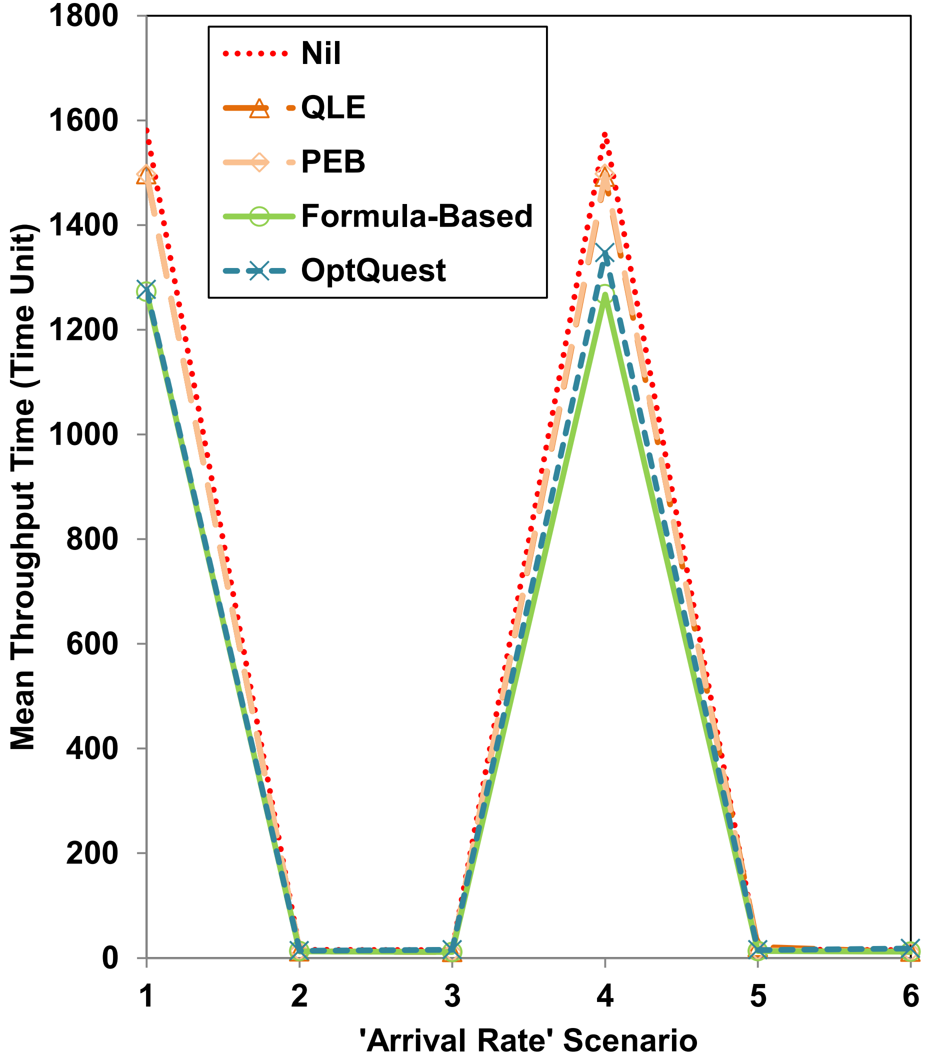


Figure 7.6: Performance Analysis on Representative Scenarios – Mean Throughput Time

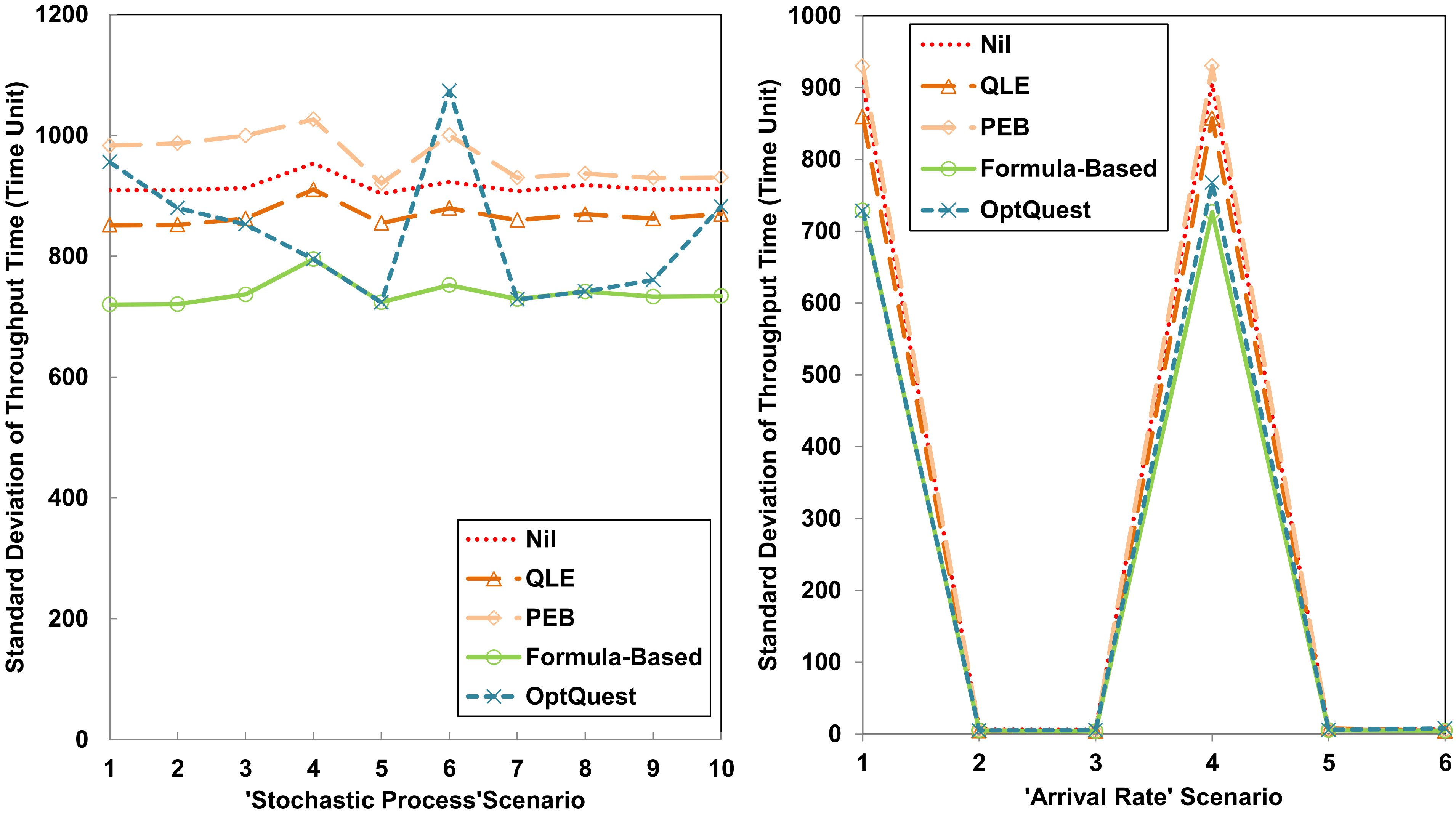


Figure 7.7: Performance Analysis on Representative Scenarios – Standard Deviation of Throughput Time

The arrival rates play a major role in the performance as shown in the right column of Figure 7.2-7. The relationship manifests itself clearly when the system is overloaded (scenarios S-7 and A-4) due to the long queue and throughput times which makes any savings crucial. Similar here, the developed method handled overloaded systems better than other methods even when c_a increased from 0 to 0.5.

As for balanced and underloaded systems (Table 7.3), queue times are very small even when the arrival rates are highly variable. In terms of the throughput rates, for balanced lines, the developed method is still recommended as it performs well, especially for the case when the intra-variability of the arrivals increases. However, with underloaded system, any of the three autonomous control methods, i.e., Formula-Based, QLE and PEB, can be used as they are all give almost the same improvement in the throughput rate.

As for simulation-based optimisation, OptQuest found the optimal solution that improves throughput and queues, as the formula-based autonomous control, on four out of the fifteen scenarios only. For these scenarios, OptQuest converged after different times (Figure 7.8).

In terms of utilisation for all scenarios, the developed formula-based autonomous control improved the utilisation efficiency rather than solely focus on increasing the utilisation. This can be seen clearly in scenarios 7, 8 and 9, where OptQuest has an increased utilisation over the developed method but the throughput and queue time remained the same for both methods and even, in some cases, lower with the developed formula-based autonomous control.

Table 7.3: Mean Queue Time and Throughput Time for Balanced and Underloaded Scenarios

No.	KPI	Control Method				
		Nil	QLE	PEB	OptQuest	Formula-Based
A2	Mean QT	0.29	0.39	0.57	0.53	0.47
	Mean TT	15.74	12.38	12.99	13.58	12.57
A3	Mean QT	0	0	0	0	0
	Mean TT	14.89	11.16	11.14	15.58	11.17
A5	Mean QT	0.45	1.42	0.94	0.76	0.61
	Mean TT	16.21	21.03	14.14	15	13
A6	Mean QT	0.05	0.14	0.22	0.26	0.16
	Mean TT	15.03	11.59	11.85	17.87	11.63

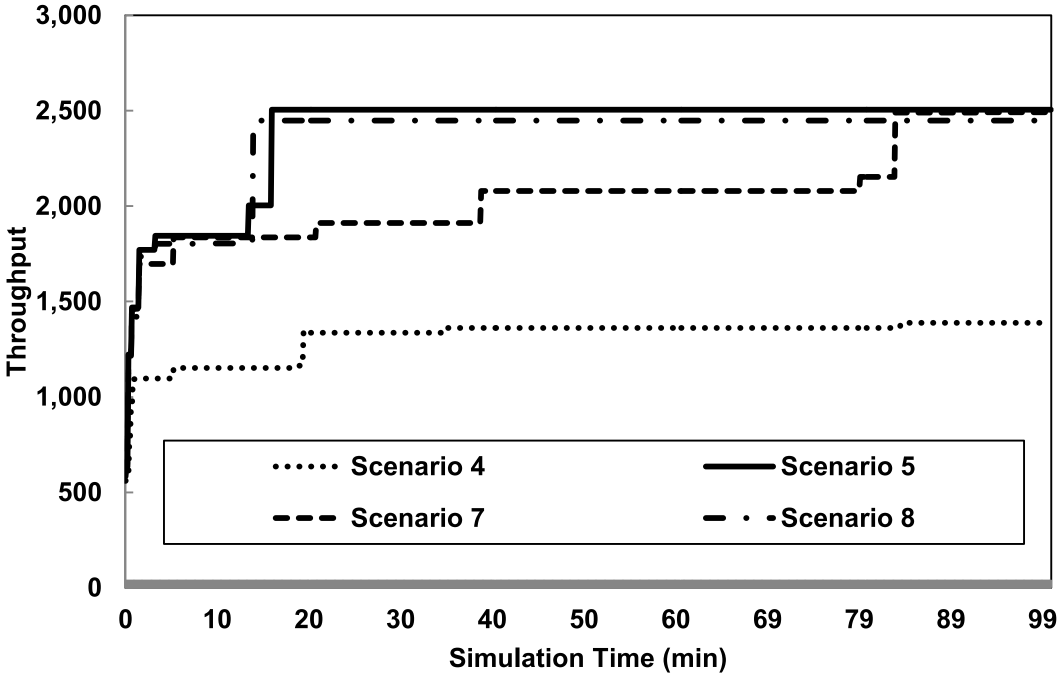


Figure 7.8: Convergence Time of Best Solution for Four Scenarios

7.3 Real-world Assessment

The idea of autonomous control of the real-world case study can be broken down into:

i. Planning Optimisation (Planning and Scheduling):
   a. Obtain the optimal arrival rate of trucks i.e. concrete delivery schedule; and
   b. Determine the expected completed barrier length and queue times based on the decision combination scenario (Table 6.18); and
ii. Operations Optimisation (React):
   a. Determine a solution, i.e., change of decision variables, to a new situation, modified decision variables or constraints, that will give the optimum impact in terms of performance. Autonomous control reacts to a change in semi-controllable factors by manipulating other factors based on the impact on the performance measure(s) of interest.

The developed autonomous control method was used to improve the process and production planning and operations of the concrete central reservation barrier (CRB) construction project for the real-world case study.

From the operational point of view, the whole idea of the CRB project improvement is to achieve a synchronous waste-free flow of material which is an essential aspect of lean operations (Dirgo 2006). The main operational objective for the industrial partner is a seamless flow of concrete loads at the construction site, i.e., maximisation of completed barrier length, i.e., throughput, and minimisation of the queues at the construction site. To achieve this goal, the arrival rate has to be adjusted to an optimal level based on the work load imposed by the variability of the system to reach to the two objectives i.e. maximum throughput and minimal queues. In other words, the production plan at the concrete plants, i.e., schedule of concrete deliveries, needs to be optimised for continuous and waste-free operation of the extruder on the other end. Furthermore, the production plan has to include decisions regarding the controllable variability factors which changes during the day or from one particular day to another. Hence, the following performance measures were measured at the end of each optimisation or autonomous control run:

i. Completed Barrier Length;
ii. Average Queue Time at the Construction Site; and
iii. Standard Deviation of Queue Time at the Construction Site.

iv. Average Queue Time at the Concrete Plant Sites; and
v. Standard Deviation of Queue Time at the Concrete Plant Sites.

In terms of the queues at the construction site, they were measured in terms of the trucks with suitable load waiting to be discharged, since the waiting generated by the load conditioning process is of the same rate from one scenario to another.

The search space of the optimisation problem was kept the same for all methods and it was limited to the realistic levels of variability that can exist for each factor as outlined in Table 6.17. The logic of the formula-based autonomous control operation is described in section 4.5.3.3. The completed barrier length, i.e., throughput, was normalised to the average current daily throughput, which need to be improved, according to the industrial partner.

Figure 7.9 shows the calculated normalised completed barrier length using the developed empirical formula, as discussed in Section 6.5, for each variability scenario with the decision combination scenario number 1 (Table 6.18). As shown, the throughput is stable with slight fluctuations up to the variability scenario number E441. After this scenario, the throughput starts to decline. Hence, the stability point E441 represents the optimal arrival rate, where to the left, queues start to accumulate and to the right, the throughput descends. This stability behaviour repeats itself for each individual decision variable. The formula-based autonomous control method used the arrival rates of the highest stability point, among all decision variables, to control the CRB project. The optimal settings from this point are then used to run the simulation model and generate the optimal delivery schedule (Appendix Q (P. A-176) gives the optimal delivery schedule for decision combination scenario number 1).

OptQuest was set with an objective to increase the completed barrier length while arrival rates were limited such that they will not produce queues at the concrete plant sites beyond the limit, i.e., 10min. Lower arrival rates were excluded from the OptQuest optimisation steps.

The previous validation step (Section 7.2) dealt with the decision problem of routing of multi products but it was limited to the case when the number of products is equal to the parallel routing. This step includes this autonomous control challenge with parallel processing that is higher than the number of products. It also deals with the challenge presented by the arrival rate equals, lower and higher than the capacity of the system.

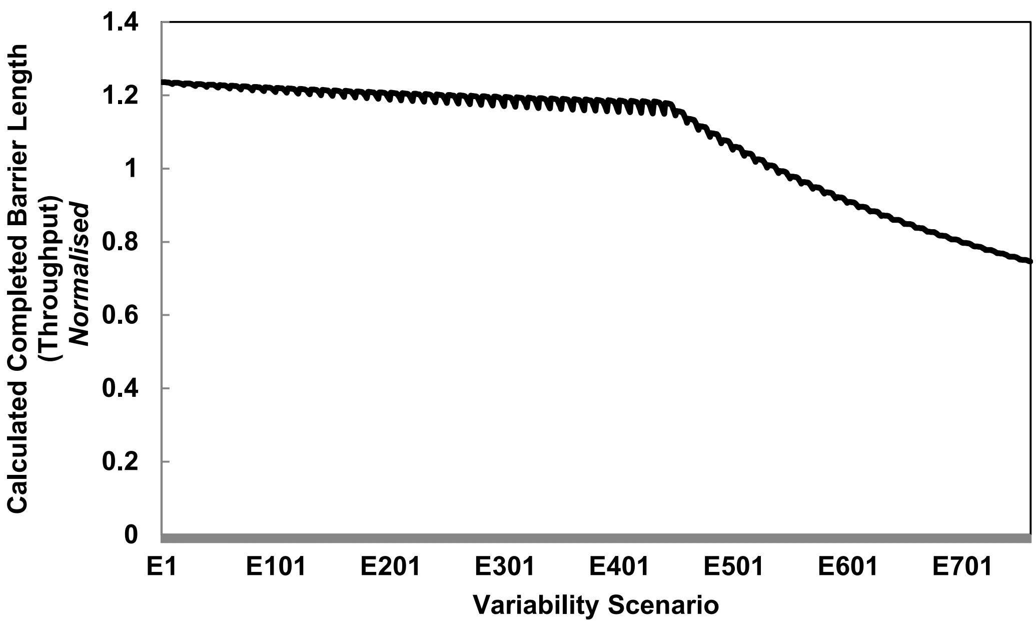


Figure 7.9: Stability (Optimal) Point for Decision Combination Scenario 1

Since the current autonomous control methods are concerned with the routing decisions in a flexible flow line setup, they could not be implemented in this case study. However, the best performing autonomous control method in the previous validation step (Section 7.2), i.e., QLE, was used here as a supplementary method to assist the formula-based autonomous control, where decisions regarding parallel processing are required, e.g., multiple concrete plant usage for batching.

The decisions here are primarily decided by the formula-based autonomous control apart from the case, where the number of parallel routes does not equal the number of products, QLE was used to take this decision.

The arrival rates for such scenarios are adjusted as follows:

i. Multiple concrete plants usage: the formula was used to calculate the optimal arrival rate for each concrete plant and the average of both was taken;
ii. Number of deliveries from the same concrete plant: if increased from one to two, the arrival rates were doubled to cope with the increased degrees of freedom in load delivery; and
iii. Both multiple concrete plants usage and double deliveries from each plant: here the arrival rate was set to the lowest optimal arrival rate, i.e., stability point, due to the increased work load on the system.

Figure 7.10-12 compares the performance of the OptQuest optimisation, formula-based autonomous control and the hybrid method of formula-based integrated with QLE autonomous control. Appendix M (P. A-178) lists the chosen operational settings for each method with the 16 decision combination scenarios.

For the decision variables not related to parallel processing, the formula-based autonomous control performed the best among the three methods with the highest throughput and minimal queues that are close to zero. However, the formula handles the parallel processing by adjusting the arrival rate to synchronise the flow of materials so that there is no need for parallel processes to reduce the work load. However, this approach does not always work since the variability of the system can generate gaps in the flow which cannot be measured by the formula since it is not applicable to parallel processes. A solution to this problem is to fill this gap with a parallel process using another autonomous control method, e.g., QLE. This side by side with the adjusted arrival rates by the formula can fill the performance gap with minimal generated queues as a result. This is evident in scenarios 3, 4 and 9 where throughput was improved with the hybrid method while queues increase, as a result, was still low.

7.4 <u>Summary</u>

The developed autonomous control method exhibited superior performance to the other autonomous control and simulation-based optimisation methods for static and dynamic situations and multi-complexity of products, i.e., inter-variability of processing times. Though the formula is only based on throughput rate, this was enough to optimise other local performance measures. However, since the formula was developed for serial flow lines only, it performed poorly in situations of parallel processing. However, this was mitigated with the integration of QLE method in to the autonomous control. The hybrid method performed well for all the variability scenarios of the real-world case study (Figure 7.10-12).

Next chapter gives a critical assessment of all the research outcomes from the Data Mining and Autonomous Control Frameworks in terms of contributions and limitations.

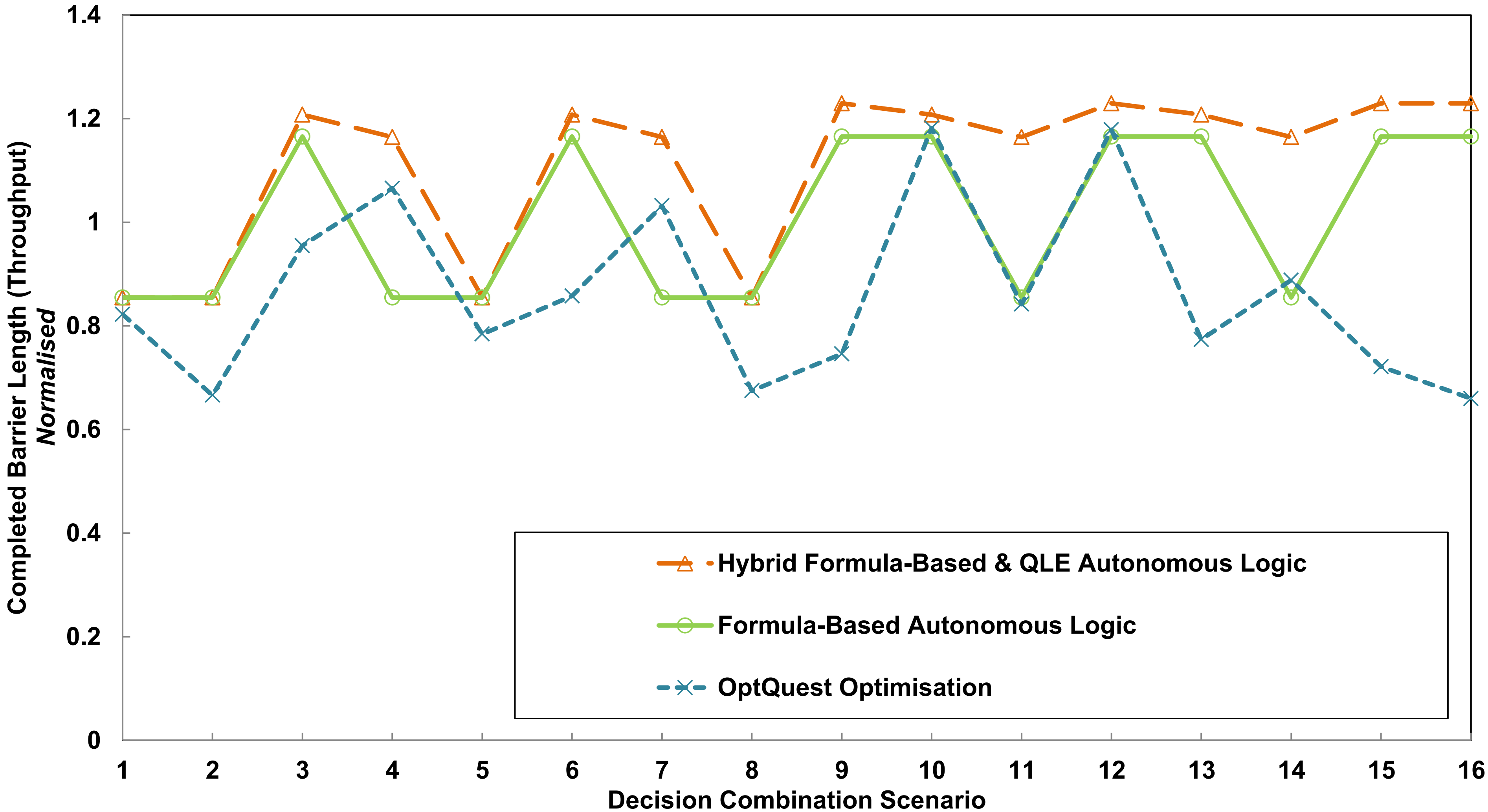


Figure 7.10: Performance Analysis on Real-world Case Study – Throughput

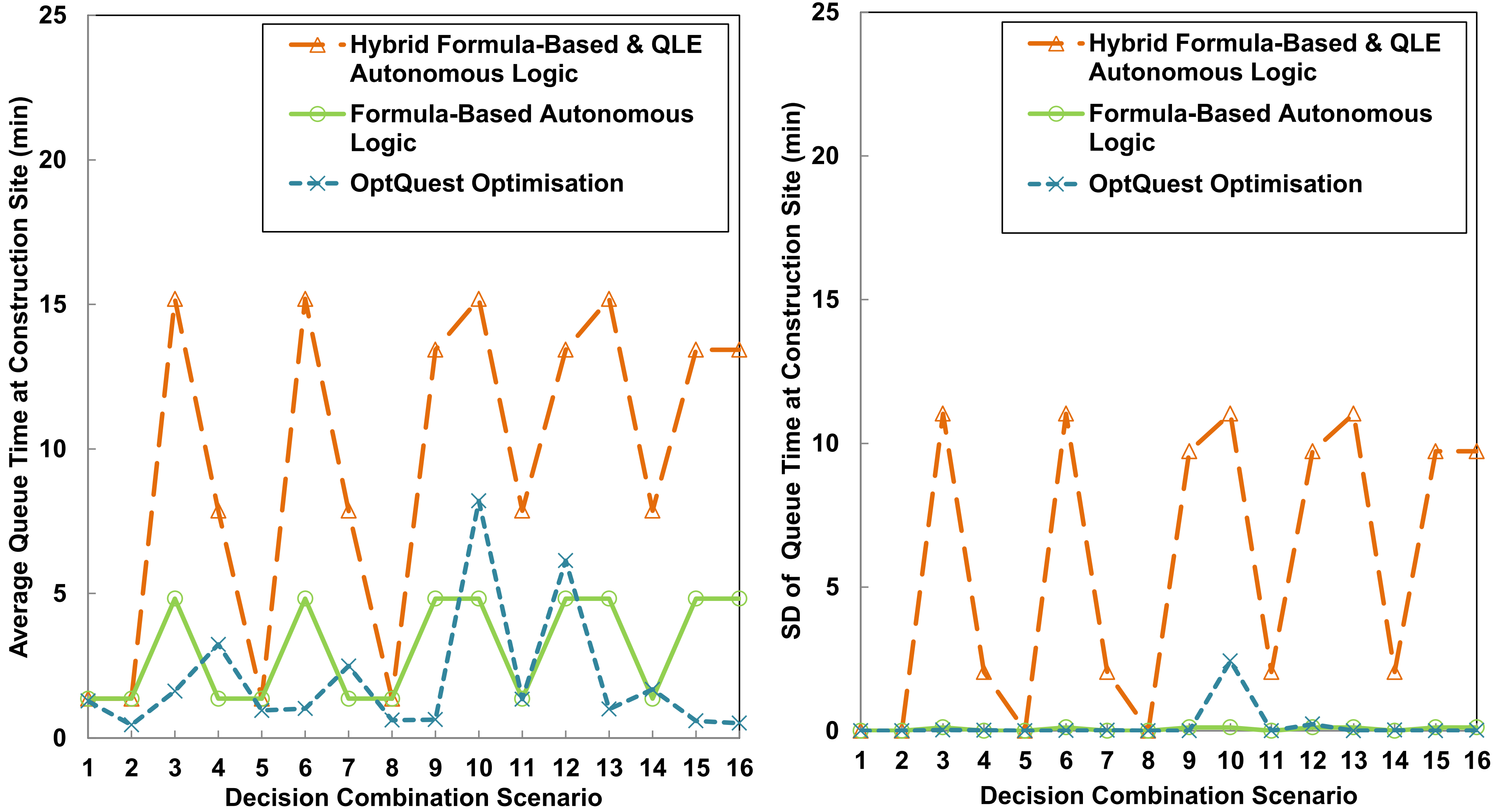


Figure 7.11: Performance Analysis on Real-world Case Study – Queue Time at Construction Site

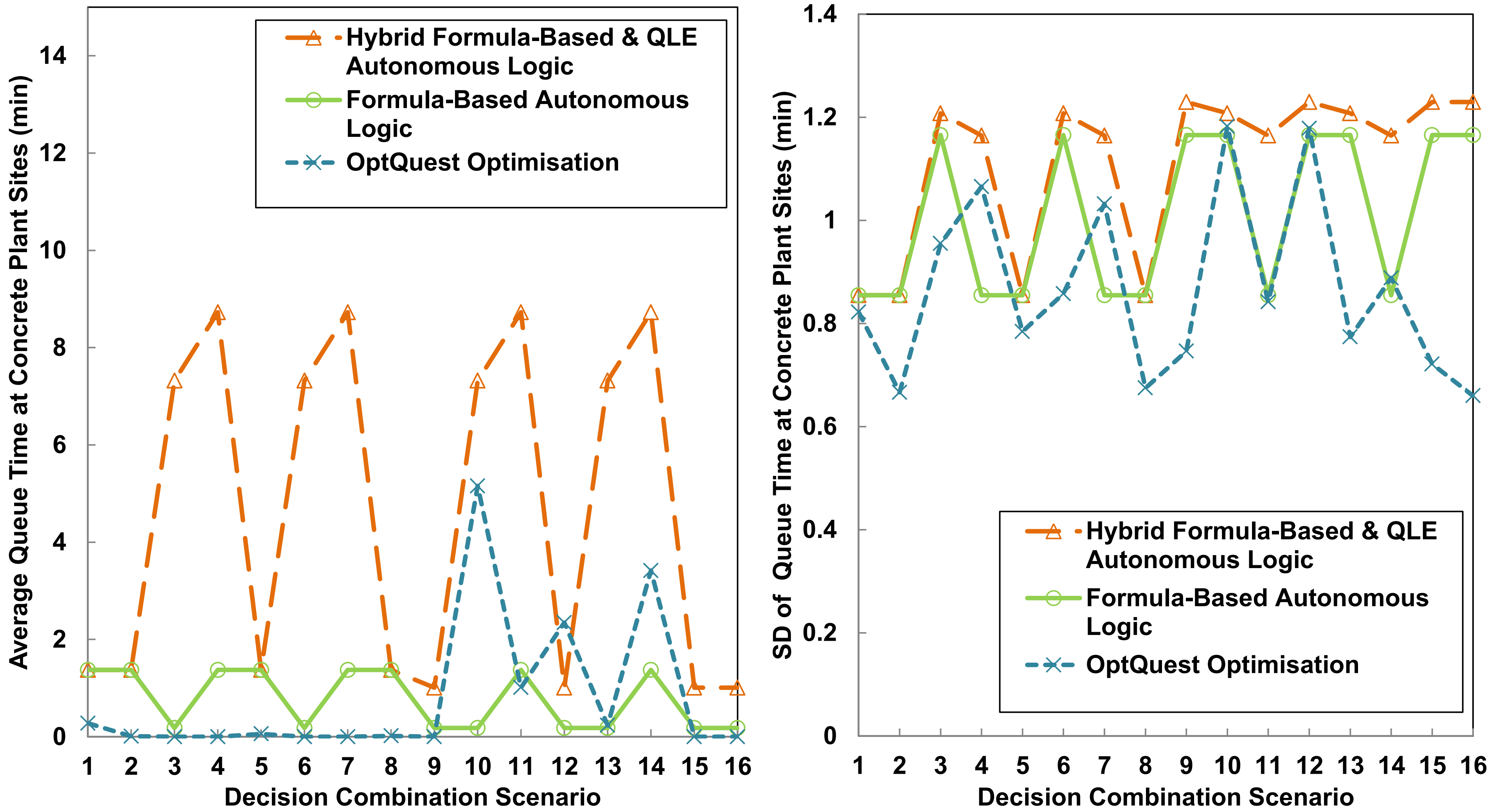


Figure 7.12: Performance Analysis on Real-world Case Study – Queue Time at Concrete Plant Sites

8 CRITICAL ASSESSMENT

8.1 Introduction

With the advancement of technology, flow lines have evolved into more intelligent and flexible structures. Flexibility represents first that the flow line is stochastic, but also that that there is a degree of controllability in the variable parameters which can be the key to control this stochasticity. Flexible flow lines have been introduced to deal with a market demand for product complexity and variety. Flexibility is a key solution to adapt with changes to the production processes. However, variable products translate to an increased variability within and across production processes, i.e., intra- and inter-process variability, which subsequently affect the performance targets. Understanding the underlying relationships between the variability within the flow line and the performance measures can be advantageous to efficiently control the variability of flexible flow lines. Fast pace of today's industrial environment requires simple and straightforward intelligent solutions to understand and act based on these relationships. This research main concern was to fulfill this requirement from both the variability modelling and control perspective for flexible flow lines.

This empirical research firstly delivered a simple standalone closed-form formula that can be used to determine the first order performance target, i.e., throughput rate, for serial flow lines with arbitrary length and stochastic normally-distributed process variability. Outcome was then used to build and validate an autonomous control method allowing for an increased throughput, improved resource efficiency and minimised queues within flexible flow lines.

The critical assessment is divided into the three main outcomes of this research:

iii. **Data Mining Framework:**
   a. Flow Line Representation;
   b. Empirical Formula For the Throughput Rate of:
      - *Synchronous Human-dependent Serial Flow Lines; and*
      - *Asynchronous Human-dependent Serial Flow Lines;*

iv. **Autonomous Control Framework:**
   a. Formula-based Autonomous Control Method; and

v. **Practical Implementation.**

This chapter discusses the research outcome in light of the main objectives and the research questions of each to:

i. examine the precision and efficiency of the results within each objective;
ii. demonstrate the contributions based on the results against the current state-of-the-art; and
iii. state the limitations.

8.2 Data Mining Framework

Data mining at its core is a toolbox of methods which can be applied to analyse pre-processed data and extract knowledge which can serve a particular research problem. A choice of which methods to pick, in the knowledge discovery, depends primarily on: suitability to the research domain, problem in hand and the step where the methods will be applied.

The main limitation of the empirical approach is that the results are not theoretically proven. This was mitigated during the definitions of data sets, testing the empirical formula and validation with a real-world case study.

8.2.1 Representation of Non-exponential Serial Flow Lines

Data are the distinguishing element of an empirical form a theoretical research (Gratton and Jones 2004). Data collection and generation were the two routes used in this research.

In terms of data collection, the study is part of a research project (InnovateUK Grant No. 18834-132285 '*Development of an innovative Autonomous Model Development Tool (AMDT) for boosting manufacturing process competencies*'), where findings of research were applied within the construction industry. Actual data were collected and validated from a concrete reservation barrier construction project on UK M1 motorway. The real-world case study resembles a flexible flow line with large-volume steady production of medium-variety products, i.e., six sizes of concrete batched from two plants.

As for the data generation, the objective here was to generate synthetic data for generic representation of the intra- and inter-process variability within synchronous and asynchronous non-exponential serial flow lines and use Discrete Event Simulation modelling to obtain the steady state simulated throughput rate with high certainty.

Several research questions were answered during the course of addressing this objective.

8.1-Q1: *Which data pre-processing method performs the best in handling the bias caused by the simulation error and increases the reliability and confidence in the simulated throughput rate?*

This research question raises an important concern regarding the reliability of the simulation results and to what extent they can be reproduced by other research to obtain the same results. Two decisions are required at this stage; one concerning how to ensure the model results are reflecting the steady state of the performance measure of interest and the second is the number of replications required to obtain the simulated throughput rate with high certainty.

For the first decision, the infinity method was chosen to reach the steady state as described in Section 4.5.2.2. Results (Section 6.2.1) show an empirical evidence that the simulation time required to reach to the state where $t_{sim} \to \infty$ for non-exponential serial flow lines is a function of two main factors; flow line length N and maximum mean processing time $\mu_{\max}$. This led to derivation of an approximation formula (Equation 6.1) for the steady state response after the defined '*Saturation Period*'. This relationship was formulated based on several experiments which showed a negligible effect of the warm-up period, i.e., variance in $TR < 1.5 * 10^{-8}$, suggesting that the simulation reached to the state, where $t_{sim} \to \infty$. This formula is limited to:

i. serial flow lines;
ii. normally distributed processes; and
iii. simulated throughput rate with 95% confidence interval.

As for the certainty, two methods were studied: Robinson (2004) algorithm to determine the number of replications required to reach a certain precision, i.e., 95% confidence interval, and smoothing of the simulated throughput rate from a single run (Section 4.5.2.3).

Results as shown in Section 6.2.1 indicate that Robinson (2004) method can increase confidence in the determined results of the throughput rate up to 7%, for the data sets used in this research. Smoothing of the simulation results, on the other hand, as demonstrated in Appendix C (P. A-6), is not a recommended approach to improve the reliability and confidence in simulated results.

8.1-Q2: *How the variability within a non-exponential serial flow line can be represented in a data mining-compatible generic form applicable to synchronous and asynchronous flow lines with arbitrary length and scenario of intra- and inter-process variability?*

The representation of the intra- and inter-process variability is an important topic since it relates to the transformation of the actual data into a meaningful generic form that is specific enough to represent the actual data yet still is generic to be applied outside the specific case study. This question was approached in this research differently from several perspectives:

i. the non-exponential distribution type that fits more for intra- and inter-human-dependent process variability within a serial flow line;
ii. representative size of the data sets, i.e., data sampling;
iii. process-based predictors, i.e., the distribution parameters used at each process i against line-based predictors, i.e., derived from the distribution parameters to represent the complete flow line; and
iv. linearity of the relationship between predictors representing the intra- and inter-process variability and the and the dependent variables, i.e., the performance target.

As described in Section 2.4, two defining parameters; effective process time and interarrival time, are generally used to describe the variability (Hopp and Spearman 2011, Hopp 2008, Etman and Rooda 2000, Jacobs et al. 2003). The effective process time includes the value-added processing time, when the process is efficiently transforming the work item, and the non-value-added times, where the work item is waiting, being in-transit, overprocessed, etc. and the interarrival time is the time between subsequent work items to arrive at a process i.

Variability can be generated within the flow line due to the following (Leu et al., 1996, Buhne et al. 2005, Hopp and Spearman 2011, Li et al. 2009):

i. machine reliability;
ii. queue capacity;
iii. natural variability leading to variations in the intra-processing time; and
iv. product complexity causing inter-processing time variation.

For human-dependent processes, the non-value-added time activities, e.g., waiting, presented in the effective process time are primarily due to the parameters three and four, i.e., intra- and inter-variability of processing times. According to Hopp and Spearman (2011) and Martin and Bridgmon (2012) variability within these two categories can be represented using the normal distribution. This agrees with the real-world industrial case study used in this research, where the industrial partner depends on the normal distribution to represent the process variability.

The data sets were generated to be representative enough for real-world synchronous and asynchronous non-exponential serial flow lines with normally distributed processes as discussed in Section 4.5.1. The data provides a wide range of flexibility for training and testing the data mining model of processing time and length variability, i.e., 1:500 units and 1:30 respectively, and continuous data (Section 5.3.1) for validation.

The third point can also be viewed from a different perspective, i.e., analytical against empirical approach to solve the research problem. From Li and Meerkov (2009), Papadopoulous et al. (2009), Li et al. (2013), Meerkov and Yan 2014, Wang et al. (2014), Kang et al. (2015) and Tan et al. (2015) it is clear that a mathematical solution requires the stochastic process to be Markovian which translates to deterministic, exponential or phase-based distributions of processing time. The attempt to solve non-Markovian process analytically (Li and Meerkov 2009, Wang et al. 2014, Kang et al. 2015) was an extension to the exponential solution which was derived through the empirical route. Looking closely at Li and Meerkov (2009) empirical formula for the throughput rate (Equation 2.8), it is a function of the minimum capacity, or the maximum mean processing time. This conclusion cannot be implied from mathematical representation of the flow line problem since this requires segmentation of the problem with a step i , i.e., mathematical solution treats each process individually despite their inter relationship to the following process.

Hence, this research opted for the empirical approach to determine the throughput rate as a function of the line-based parameters. Papadopoulos et al. (2009) suggested that future analysis of flow lines will be driven by an integration of both analytical and simulation modelling. This study proposes the combination of simulation modelling and data mining-based methods as a potential alternative.

This research used the main line-based parameters that distinguish the throughput rate of exponential from non-exponential asynchronous flow lines, i.e., the maximum mean processing time $\mu_{\max}$, the average coefficient of variation c_{av} and length N.

It also added there new predictors inspired from these parameters for investigation purpose:

i. minimum mean processing time $\mu_{\min}$;
ii. average mean processing time μ; and
iii. coefficient of variation of mean processing time c.

Furthermore, due to the fact that the process with the maximum mean processing time plays a major role, it was worth investigating if its location within the flow line has any effect on the throughput rate.

Hence, for a more generic representation of the asynchronous non-exponential flow line, the following flow line-based parameters were investigated during this empirical research:

i. Minimum Mean Processing Time $\mu_{\min}$;
ii. Average Mean Processing Time μ;
iii. Maximum Mean Processing Time $\mu_{\max}$;
iv. Coefficient of Variation of Mean Processing Time c;
v. Average Coefficient of Variation c_{av};
vi. Location Ratio of the Process with Maximum Mean Processing Time l; and
vii. Length N.

These reduce down to the main process-based parameters, μ_i, c_i and N, for synchronous flow lines.

Furthermore, looking at the closed-form formulas for synchronous flow lines built by Muth (1987) and Blumenfeld (1990), the data mining models were based on multiple linear regression of linear terms. Blumenfeld and Li (2005) used the analytical approach to produce a closed-form formula for the throughput rate of synchronous flow lines, however, as discussed this required the processing times to be deterministic and exponentially distributed interruptions. Li and Meerkov (2009), Wang et al. (2014), Meerkov and Yan (2014) and Kang et al. (2015), on the other hand, applied approximation methods to Markovian analysis of an asynchronous flow line with stochastic Markovian processes. Afterwards, they extended the analysis to include the non-Markovian case using the empirical approach. However, in this extension, they also depended on the linear terms of the line-based parameters. It is clear in Li and Meerkov (2009) formula that the process variability, represented by the maximum mean processing time $\mu_{\max}$, the average coefficient of variation c_{av} and length N, reduces the throughput rate from the exponential case. However, the linearity assumption can cause the relationship between the variability parameters and the performance targets to not be fully established. This research targeted this issue by investigating if a nonlinear relationship between the variability parameters, i.e., predictors, and the performance target, i.e., dependent variable, exists.

Statistical analysis on the relationship between these parameters including their nonlinear terms and the throughput rate was carried out as discussed in Section 6.3. Based on the results, new nonlinear relationships between the following set of flow line-based variability parameter terms and the throughput rate were confirmed to a high certainty as follows:

i. The inverse of maximum mean processing time $\mu_{\max}{}^{-1}$, the coefficient of variation c_{av} and N in agreement with the literature;
ii. Additionally, a nonlinear term related to the coefficient of variation, namely $e^{c_{av}}$; and
iii. Furthermore, one term for the length, i.e., N^{-1}, was also identified.

Results also showed that although the process with maximum mean processing time has a significant effect on the throughput rate, the location of such process is irrelevant. Furthermore, parameter terms $e^{\mu_{\min}^{-1}}$, μ , μ^{-1} , $\log \mu$, $e^{\mu^{-1}}$, c , c^{-1} and $\log c$ with inconsistent relationship with the throughput rate but an acceptable statistical importance suggesting that a relationship might exist.

Best regression technique was applied to verify the findings. Linear regression model with the first set of parameter terms was accurate to a standard error of 0.0082 and R^2 of 96.7%. Accuracy was slightly improved with the second set were added to R^2 of 97.3% and 0.00686 standard error.

Identification of the parameter terms in this manner allows for implementation of multiple nonlinear regression without prior knowledge of the model expression by using multiple linear regression of nonlinear terms.

In conclusion, the main contribution of this section is a generic representation of synchronous and asynchronous non-exponential flow lines with arbitrary length and intra- and inter-process variability. Generic flow line-based, instead of process-based parameters, were used to target the parameters with impact on the throughput rate and allow for a simple closed-form formula to be developed for the relationship between variability parameters and the throughput rate. Results show that the parameters can be categorised into two categories. The first one consists of the main nonlinear parameter terms with clear impact on the throughput rate are $\mu_{\max}{}^{-1}$, c_{av}, $e^{c_{av}}$, N and N^{-1}. The second category includes parameter terms with less statistically proven relationship with throughput rate. These terms were included as free predictors with the intention of using them to improve the accuracy of the model; these free predictors are $e^{\mu_{\min}^{-1}}$, μ , μ^{-1} , $\log \mu$, $e^{\mu^{-1}}$, c , c^{-1} and $\log c$.

8.2.2 Empirical Closed-form Formula for the Throughput Rate of Human-dependent Serial Flow lines

Several flow line models were developed using different techniques. Earlier models gave exact mathematical models but were limited to short flow lines. Recent models used approximate analytical solutions, such as decomposition and aggregation methods, to model arbitrary length flow lines. However, these models still require the process to be Markovian, hence, they cannot be applied to human-dependent processes with normally distributed variability patterns.

Simulation modelling of flow lines has developed significantly giving an accurate approximation of actual flow lines. However, simulation models do not formulate the relationships within the system and is usually used to represent a specific case study with limited applicability for other case studies without major modifications to the simulation model itself.

Several researchers have investigated development of closed-form formulas for quick analysis of a wider range of case studies of exponential and non-exponential flow lines at the cost of applying some simplifications or assumptions to the flow line. closed-form formula also gives an added advantage to the analysis related to process and production planning since it can be easily interpreted and implemented in industry. Closed-form formulas provide an easy and time-efficient approach to evaluate variability which is advantageous for a quick autonomous decision. The objective here was to build a Data Mining Framework and use it to develop a standalone empirical formula and perform goodness-of-fit analysis for the estimated throughput rate for synchronous and asynchronous human-dependent serial flow lines.

Development in this area started by building a standardised supervised machine learning data mining approach '*Degree of Freedom (DOF)*' based on the degrees of freedom at each step of the Data Mining Framework to build the empirical formula. This was then translated into a MATLAB program following the procedure shown in Figure D.1 and D.2 for synchronous and asynchronous flow lines respectively.

Throughout the performance analysis stage, cross validation was applied to each individual data set $\mathbf{T}_x$ within the test set $\mathbf{T}$ and to determine the goodness of each individual set using the method m. The mechanism of partitioning during the cross validation is sequential forward selection of the training sets. The cross validation was applied here to ensure that the no under- or over-fitting take place and that the pattern of the relationship between the predictor terms and the throughput rate is accurately modelled.

During the first step, the best performing data set is elected to the second step, where the remaining data sets are compared to elect a data set based on goodness of fit of each individual data set within the test set and so on. In each modelling step, the best regression model was extracted based on the minimal Mean Absolute Percentage Error (MAPE) and the stability of error across the different data sets included within both the training and test sets.

8.2.2.1 Synchronous Flow Lines

8.1-Q3a: *Can data mining models produce a simple closed-form formula to estimate the throughput rate of synchronous human-dependent serial flow lines? How accurate will this evaluative model perform for the real-world case study?*

Several analytical and empirical-based evaluative models were developed for the exponential and non-exponential synchronous flow lines covering wide range of performance measures. A closed-form formula for the steady state throughput rate of synchronous serial flow lines with normal-distributed processes was developed by Blumenfeld (1990). The '*Degree of Freedom (DOF)*' approach was implemented here to investigate if the accuracy of this formula can be improved.

The contribution of this section is the development of a simple standalone fifth-degree polynomial empirical formula for the steady state throughput rate of synchronous human-dependent serial flow line with arbitrary length and intra-process variability, with improved estimation accuracy over the latest development in this area, i.e., μ_e of 0.2% with a stability over the range of test data sets, i.e., $c_e = 0.11$ over μ_e and c_e, of 2.63% and 0.74 for Blumenfeld (1990) formula. The limitation here that this formula was not validated outside the defined synthetic discrete data sets since the continuous data of the case study represents an asynchronous flexible flow line.

8.2.2.2 Asynchronous Flow Lines

8.1-Q3b: *Can data mining models produce a simple closed-form formula to estimate the throughput rate of asynchronous human-dependent serial flow lines? How accurate will this evaluative model perform for the real-world case study?*

Papadopoulos (1996) investigated a generalisation of a closed-form formula that can fit arbitrary asynchronous flow line length of exponential distributed processes with only one distribution parameter, i.e., μ. However, the resulting formula included coefficients that need to be numerically obtained. Empirical approach, centered on simulation modelling, was the most popular for the non-exponential case, where, at present, theoretical proof is not possible (Li and Meerkov 2005, Li and Meerkov 2009, Wang et al. 2014, Kang et al. 2015).

Li and Meerkov (2009) developed an empirical formula as an extension to the exponential case which has been the cornerstone for several other implementations including service based flow lines (Wang et al. 2014).

Current published work on this subject gives good insights and ideas, however, the formula still remains linked to the exponential case. An interesting implementation of the Markovian analysis presented in Wang et al. (2014) and Kang et al. (2015) led to a closed-form expression for the exponentially-distributed throughput that does not require numerical iteration. In Wang et al. (2014), the state space was used to represent the number of patients in each treatment stage, where resources, e.g., nurse, are allocated based on their availability. The processing time of each resource is exponentially distributed and it was considered the same for all resources of the same resource group. The results of Wang et al. (2014) are then adjusted using the empirical formula of Li and Meerkov (2009) for non-exponential processes. In this research, due to the limitations of the Markovian analysis listed in Section 2.6, which are still applicable to this case, and also for improved accuracy, the main benchmark, i.e., simulation, was used to determine the throughput rate of the exponential case, i.e., $c_{av} = 1$, and then used to feed Li and Meerkov (2009) formula.

The contribution of this section is the development of a simple standalone third-degree polynomial empirical formula for the steady state throughput rate of human-dependent asynchronous serial flow line with arbitrary length and intra- and inter-process variability, with improved estimation accuracy over the latest development in this area, i.e., μ_e of 2% with a stability over the range of test data sets, i.e., $c_e = 0.19$ over μ_e and c_e, of 2% and 0.45 for Li and Meerkov (2009) formula with simulation used to determine TR^e. The developed closed-form formula also removes the dependency of the formula on external inputs, e.g., from simulation or analytical solution, by producing a higher order formula that includes all generic variability parameters contributing to the throughput rate.

It is worth noting that a single nonlinear predictor term from each variability parameter, i.e., ${\mu_{\max}}^{-1}$, $e^{c_{av}}$, N^{-1} and $e^{\mu^{-1}}$, was sufficient to form the relationship with the performance target, i.e., throughput rate.

8.3 Autonomous Control Framework

Autonomous systems, in its general concept, are designed to process available and up-to-date information at the autonomous entity level about current state, including local performance measures, and use them to tweak the decision variables to achieve a specific objective. This approach addresses decentralisation of decision and responsiveness in a dynamic environment. Optimisation, instead of depending on the performance measures of the current state, gradually learns how variability of the system affects the objective system-level performance measure(s) and adjusts the decision variables accordingly. The objective here is to build an Autonomous Control Framework for flexible flow lines based on the developed empirical formula and compare it with current autonomous control and simulation-based optimisation methods using representative variability scenarios of flexible flow lines.

8.3.1 Formula-Based Autonomous Control Method

8.2-Q1: *Can this formula-based evaluative model be utilised to control the variability within a flexible flow line?*

Yes, the developed formula was used as an autonomous-decision-support system by investigating the link between autonomous decisions according to the manufacturing flexibility and the overall performance targets. The empirical formula estimates the throughput rate at the current state and chooses the autonomous decision step(s) that increases the throughput rate.

8.2-Q2: *If yes, how does it compare to other existing methods in terms of performance?*

Intelligent and autonomous systems rely solely or partially on autonomous control and scheduling optimisation heuristics for a rapid response to the daily challenges faced by production and process planners. Autonomous methods respond quicker to sudden changes in the system while optimisation usually requires more time to learn from past events. However, since optimisation can target system-level performance measures, such as in this case, it can theoretically reach to a better solution than QLE and PEB autonomous control methods. Hence, in Scholz-Reiter et al. (2010), autonomous control usually outperformed optimisation in a dynamic environment while scheduling heuristics were most of the time the best option during static scenarios.

This research presented a predictive autonomous control, where the decision is still in the hands of each individual process as in the autonomous control. However, instead of depending only on local information at one or more processes, the production planning decision here, and accordingly scheduling, is rendered based on a closed-form formula that predicts the outcome of variability on the system-level performance targets, hence, increased optimisation certainty.

Scholz-Reiter et al. (2005) model of flexible flow line was the basis for implementation of many of the autonomous control methods (Scholz-Reiter et al. 2006, Scholz-Reiter et al. 2008a, Scholz-Reiter et al. 2008b, de Beer 2008, Scholz-Reiter et al. 2010, Windt et al. 2010, Grundstein et al. 2015). The model was reproduced here and compared with the reported results for QLE method in Scholz-Reiter et al. (2005) to ensure that the model and method are implemented correctly.

As discussed in Section 7.2, the developed formula-based autonomous control reached to the optimum solution with static and dynamic scenarios. In a dynamic setup, QLE autonomous control logic followed the developed autonomous control method for overloaded system scenarios with synchronous change in coefficient of variation c of the processing times in agreement with the literature. However, as expected since QLE does not take c into consideration, QLE method performance degrades with the increase in the synchronous change in coefficient of variation giving an advantage to the optimisation to exceed the QLE performance. For the asynchronous scenarios, QLE performance drops further to reach a performance slightly higher than the PEB method. On the other hand, all autonomous control methods including the developed formula-based performed nearly the same in underloaded dynamic situations.

As for the balanced static scenario, OptQuest optimisation was the best performing after the developed formula-based autonomous control.

The QLE method provides quick response and adaptability to sudden changes in customer order frequency in the flexible flow line. However, it also has some limitations as follows:

i. it cannot handle the changes in coefficient of variation of processing times;
ii. queue length has to be known beforehand to apply the control logic;
iii. suitable for flow lines with queues only, i.e., if no queue exists within the flow line, or part of it, the technique will choose the routing with smaller processing time regardless of any blocking or starvation that might occur as a result; and
iv. inter-relationship between processes are not determined; therefore, it cannot detect problems in the flow line due to unsynchronised processing times from one process to the succeeding one.

The contribution of the Autonomous Control Framework is a formula-based autonomous control method for production planners to determine the optimal operational settings based on the constraints imposed on the system by relating the variability factors to the corresponding process parameters of the flow line and using the formula to obtain the optimal decision with higher certainty than current autonomous control and simulation-based optimisation methods. The formula-based autonomous control method developed here has a built-in formulated relationship between the independent variable and system-level performance target, i.e., throughput rate. Hence, this improves the reliability and certainty of the autonomous decision while providing faster reaction to changes in the system at the same pace as the autonomous control. The main limitations of the technique that it can handle flow lines with normally distributed processing and interarrival times only. For routing decision, it is also limited to the case that the number of products matches the number of available flexible processes at each decision step, however, this was accounted for as reported in section 7.3.

8.4 Industrial Implementation

This research is part of the Innovate-UK project (Grant No. 18834-132285 '*Development of an innovative Autonomous Model Development Tool (AMDT) for boosting manufacturing process competencies*'). This gave the research the privilege to validate the research outcomes and findings on a real-world case study. The objective here was to validate the developed empirical formula and autonomous control method using a real-world case study in the construction industry.

8.3-Q1 *How accurate the developed empirical formula evaluative model performs for the real-world case study?*

The formula was developed using discrete synthetic data sets. The real-world case study allowed testing the formula on continuous data.

The results described in Section 6.5 show that the error increased from MAPE of 2% with discrete data to 5% when tested on continuous data, and the relationship between the changes in errors and the variations in mean processing time and coefficient of variation is insignificant. Hence, the formula still remains competent and it is expected to still provide accurate estimations of the throughput rate when applied for other case studies with different continuous data.

8.3-Q2 *How developed formula-based autonomous control method compares to other existing methods in terms of performance in a real-world setup?*

The real-world case study allowed testing the formula-based autonomous control method for other autonomous decisions apart from routing, such as in the 3x3 flexible flow line model case. On the other hand, this added a limitation on the methods that can be used in the validation. While the formula-based autonomous control and optimisation are capable to perform a decision on several dimensions of the manufacturing flexibility mentioned in Section 3.2, autonomous methods are primarily dedicated to the routing problem.

Hence, results as reported in Section 7.3, compare the developed autonomous control method to the OptQuest optimisation. The decision variables cover a wide range of flexibility including process, routing and product decisions.

The main objective here was to determine primarily the normal distribution parameters of the interarrival time and secondly optimisation decisions on the variability parameters that can balance the system and provide the optimum throughput with minimal queues. The secondary decisions handle a standard optimisation problem, where an optimal setting of variability factors is pursued. This includes the case, where the number of routes is higher than the number of products, i.e., parallel processing. The developed method outperforms OptQuest on many levels apart from the parallel processing scenarios. Hence, QLE autonomous control method, giving its high performance as reported in Section 3.5.2 and 7.2, was investigated here as a supplementary method to the developed formula-based autonomous control method. This subsequently led to another contribution presented by the development of a hybrid autonomous control method integrating the developed autonomous control method and QLE. The hybrid autonomous control method was able to perform and give the optimal solution where the formula-based autonomous control failed alone to achieve. Specifically, in parallel processing, the hybrid autonomous control method allowed releasing the pressure on overloaded processes which subsequently increased the throughput rate, where the formula cannot predict this since it is not designed for parallel processing.

A real-world trial was carried out as part of the Innovate-UK project (Grant No. 18834-132285 '*Development of an innovative Autonomous Model Development Tool (AMDT) for boosting manufacturing process competencies*'), where Full factorial DOE combined with Discrete Event Simulation (DOE-DES) was used to obtain the best operational settings that will maximise the throughput rate and minimise queues at the construction site. The DOE-DES was used to run through all the possible scenarios and then the optimum scenario was chosen. The primary objective was maximising the throughput, i.e., completed barrier length, and as a secondary objective to minimise queues at the construction site with a margin of 10%. The same range of decision variables, as shown in Table 6.17, was used with the DOE-DES apart from the mean and coefficient of variation of the arrival rate to reduce the number of scenarios. The range of the mean arrival rate in DOE-DES was kept at 10, 40 and 80 trucks per min while the coefficient of variation was set to 0. Another difference in the model is that with DOE-DES the trucks were allowed to return back to the concrete plant in a closed loop to reduce the overall truck rental cost.

This produces disruptions to the arrival rate, which was not taken into account in this research to investigate the optimum arrival rate more accurately. The trial followed the E14 scenario presented in Table 6.18 and the results of the DOE-DES implemented during the real-world case study compared to the methods reported here is shown in Table 8.1. Results clearly shows that both DOE-DES and the optimal method, i.e., developed hybrid autonomous control method, gave the same decisions for all decision variables except the arrival rate, where the DOE-DES is limited to three levels only. DOE-DES, however, chose the closest value to the optimum from its range, i.e., 40/min. It has to be noted that the arrival rates for DOE-DES are ambiguous due to the feedback from the construction site to the concrete plants. The hybrid method took advantage of its wider range to adjust the arrival range to increase the throughput and reduce queues at the construction site. The hybrid method gave a performance increase of 20% in the throughput in comparison to DOE-DES. Additionally, mean queue times were minimised by 23min, however, more variable, i.e., ,standard deviation of queue times increased by 7min in total, due to the increased number of trucks within the system.

The findings from Section 7.2 and 7.3 reveal the potential of the developed hybrid autonomous control as a competitive autonomous control method that can accommodate for several manufacturing flexibility levels and provides more optimisation certainty. The main limitation of the method is that it requires the processes of the flow line to be normally distributed. It was also not tested for the decision problem case, where the number of products exceeds the number of routes.

8.5 Summary

The outcomes of this research showed contributions to the current state-o*f*-the-art in the field of evaluative modelling and autonomous control of flow lines. The research benefited also from being part of an InnovateUK research project, to validate the outcome on a real-world industrial case study which showed promising results. Next chapter summarises this research and gives recommendations on potential future steps.

Table 8.1: Comparison of the Trial Results to the Developed Formula-based and Hybrid Autonomous Control (AC) Methods and OptQuest Optimisation

Method	No of Deliveries from the Same Concrete Plant	Concrete Plant	Load Size (m3)	Multiple Concrete Plant Usage	Arrival Rate (1/min)		Mean Site Delay (min)	Throughput (normalised)	QT at Construction Site (min)	
					Mean	Standard Deviation			Mean	Standard Deviation
DOE-DES (Trial)	2	2	8	S	40	0	0	0.974	31.94	4.59
Formula-based AC	1	2	8	S	49	0	0	0.855	1.37	1.45
Hybrid AC	2	2	8	S	25	0	0	1.164	8.73	11.31
OptQuest	1	2	8	S	28	2.6	0	0.888	50.42	40.03

Decision Variable

9 CONCLUSIONS AND FUTURE WORK

9.1 Conclusions

The 21st Century competitive market demanded for changes in the manufacturing and production culture to become more flexible and lean to satisfy the 21st Century customer needs. Customised complex innovative products enter the market on short intervals to meet the market demand. This comes at the price of increasing the variability within the process and production planning and scheduling of the flexible flow lines producing these products. In lean production, control of the effects of variability within the process and production planning and scheduling on performance targets is a strategic goal.

Autonomous control systems decentralise the control strategy of flow lines by handing over the process and production planning and scheduling decision to the intelligent entities, i.e., processes. This approach usually outperforms centralised heuristic scheduling in a dynamic setup while the first works better for static situations. Autonomous systems can make use of the development within evaluative modelling for a more stable performance regardless of the system dynamics.

This research investigated development of an autonomous-decision-support closed-form formula through empirical evaluative modelling that can help to easily and quickly estimate the effect of each autonomous decision, based on the stochastic variability in process and production planning, on the system-level performance of flexible human-dependent serial flow lines.

Through this investigation, the following main contributions to knowledge were achieved:

i. Generic representation of arbitrary length non-exponential serial flow line using nonlinear terms (Section 4.3). New nonlinear relationships between the normal distribution-based variability parameters and the throughput rate were identified with p-values less than 0.01 and correlation coefficients higher than 0.8. Namely, ${\mu_{\max}}^{-1}$, c_{av}, $e^{c_{av}}$, N and N^{-1} (Section 6.3); and

ii. Simple standalone closed-form empirical formulas that estimate the throughput rate of synchronous and asynchronous flow lines with normally distributed process variability to a higher accuracy and independency than currently available formulas.

The polynomial stepwise regression model with bounded steps for synchronous flow lines (Equation 6.3) gave an average prediction percentage error μ_e of 0.2% with a stability over the range of test data sets, i.e., $c_e = 0.11$. This model surpasses the performance of Blumenfeld (1990) formula which gave, for the same training data set, an average and coefficient of variation of the prediction percentage errors, μ_e and c_e, of 2.63% and 0.74 respectively. As for asynchronous flow lines, the best performing standalone regression model with the optimum training set (Equation 6.4) was able to give the same μ_e of 2% as the non-standalone Li and Meerkov (2009) formula with an improved c_e of 0.19 against 0.45 for Li and Meerkov (2009) formula.

iii. Formula-based autonomous control method and a hybrid formula-based and QLE autonomous control method for production planners to determine the optimal operational settings based on the constraints imposed on the system by relating the variability factors to the corresponding process parameters of the flexible flow line and using the formula to obtain the optimal decision with higher certainty than current autonomous control and simulation-based optimisation methods. For the 3x3 flexible flow line described in Scholz-Reiter et al. (2005), the developed technique outperformed QLE, PEB autonomous control and OptQuest optimisation in terms of throughput rate, queue time and utilisation efficiency against all the identified synchronous and asynchronous flow line scenarios and underloaded, balanced and overloaded systems (Section 7.2). As for the real-world case study, for the decision variables not related to parallel processing, the formula-based autonomous control performed the best among the three methods, i.e., formula-based and hybrid formula-based and QLE autonomous control and OptQuest optimisation, with the highest throughput and minimal queues that are close to zero (Section 7.3). For parallel processing scenarios, the hybrid formula-based and QLE autonomous control managed to bring the throughput to the optimal levels, however, this came at the cost of increased queues but within limits.

9.2 Future Work

The contribution of this research is represented into three main research outcomes: a generic representation of the variability of non-exponential serial flow lines, an empirical closed-form formulas for the throughput rate of synchronous and asynchronous human-dependent serial flow lines and a formula-based and hybrid autonomous control methods for human-dependent flexible flow lines. Research in these areas can be extended through the following future work:

i. Development of a graphical user interface that can be used for end-users to estimate the throughput rate based on data entry of process information. The user interface can be linked to simulation model for direct control of the flexible flow line.
ii. Expand the generic representation of flow lines to take into account the parallel processing which might occur at one or more processing stages throughout the flow line;
iii. Investigate how the developed empirical formula for asynchronous serial flow lines with normally distribution process variability performs for other exponential and non-exponential distributions such as Erlang, gamma, uniform, triangular, binomial and lognormal distributions and what extensions or improvements are required in this matter;
iv. The Data Mining Framework presented in this thesis can be applied to generate a generic representation and an empirical formula for the throughput rate of machine-dependent flow lines with medium to high Markovian process variability, i.e., short and long interruptions due to machine setup, failure and repairs. This would be particularly useful given the amount of interest in this area and the several exact mathematical solutions and approximation methods developed for this type of flow lines which can serve as benchmarks for the outcome;
v. Study the use of the developed Data Mining Framework for inclusion of finite queues into the human- and machine-dependent flow lines and generate a data mining model of the throughput rate that adds the queue capacity and its nonlinear terms as independent variables. This study can target both synchronous and asynchronous flow lines;

vi. An interesting point of research is to extend the Data Mining Framework to handle a flow line with mixed Markovian and non-Markovian processes. This can be useful to a supply chain, where manufacturing is a stage of the chain that also includes other human-dependent processes, e.g., paper work, delivery, etc.;

vii. Study the performance of the developed Autonomous Control Framework and the effectiveness of integration of more methods to handle increased production planning and scheduling constraints, e.g., due date, and higher degrees of process complexity, e.g., dynamic variability; and

viii. Investigate the application of both data mining and Autonomous Control Frameworks into more specialised versions of flow lines, than the general serial open-loop flow line case, such as assembly lines, flow line with rework or closed-loop systems. This can be particularly useful for the case study of this research, where concrete trucks can run in a closed loop by returning to the concrete plant for re-batch, i.e., multiple deliveries per a concrete truck, instead of using a new truck per delivery. This can reduce the number of work items inside the flow line which might help to improve the system performance.

REFERENCES

Ambani, S. M. (2011). *Analytical estimation of throughput distribution for serial manufacturing systems with multi-state machines and its application*. PhD diss., Michigan: University of Michigan.

Anderson, D. and Moodie, C. (1969). Optimal Buffer Storage Capacity in Flow line Systems. *International Journal of Production Research*, 7: 233-240.

Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics*, 16: 523-531.

Angeli, C. (2010). Diagnostic Expert Systems: From Expert's Knowledge to Real-Time Systems. *Advanced Knowledge Based Systems: Model, Applications & Research*, 1: 50-73.

Armbruster, D., de Beer, C., Freitag, M., Jagalski, T. and Ringhofer, C. (2006). Autonomous control of production networks using a pheromone approach. *Physica A: Statistical Mechanics and its Applications*, 363(1): 104-114.

Baird, G. (2010). *Sustainable Buildings in Practice: What the Users Think*. London: Routledge.

Banks, J., Carson II, J. S., Nelson, B. L. and Nicol, D. M. (2005). Discrete-event system simulation. 4th ed., New Jersey: Pearson Prentice Hall.

Barten, K. (1962). A Queuing Simulator for Determining Optimum Inventory Levels in a Sequential Process. *Journal of Industrial Engineering*, 13: 245-252.

Basu, R. (2009). *Implementing Six Sigma and Lean*. Oxford: Elsevier.

Bauer, A., Browne, J., Bowden, R. and Duggan, J. (1994). *Shop Floor Control Systems: From design to implementation*. Dordrecht: Springer Netherlands.

Baykasoglu, A. and Gocken, M. (2011). A simulation based approach to analyse the effects of job release on the performance of a multi-stage job-shop with processing flexibility. *International Journal of Production Research*, 49(2): 585-610.

Beaty, R. T. (1996). Mass Customization. *Manufacturing Engineering*, 75(5): 217- 220.

Benitez, J. M., Castro, J. L. and Requena, I. (1997). Are Artificial Neural Networks Black Boxes?. *IEEE Transactions on Neural Networks*, 8(5): 1156-1164.

Bley, H., Reinhart, G., Seliger, G., Bernardi, M. and Korne, T. (2004). Appropriate Human Involvement in Assembly and Disassembly. *CIRP Annals*, 53(2): 487-558.

Blumenfeld, D. E. (1990). A simple formula for estimating throughput of serial flow lines with variable processing times and limited capacity. *International Journal of Production Research*, 28 (6): 1163-1182.

Blumenfeld, D. E. and Li, J. (2005). An Analytical Formula for Throughput of a Production Line with Identical Stations and Random Failures. *Mathematical Problems in Engineering*, 3: 293-308.

Box, G. and Draper, N. (1987). *Empirical Model-Building and Response Surfaces*. Minnesota: Wiley.

Brandimarte, P. and Villa, A. (1999). *Modelling Manufacturing Systems: From Aggregate Planning to Real-Time Control*. Berlin: Springer-Verlag.

Brooks, R. (1987). Intelligence without representation. *Artificial Intelligence*, 47: 139-159.

Buhne, S., Lauenroth, K. and Pohl, K. (2005). Modelling Requirements Variability across Product Lines. *Proceedings of the 2005 13th IEEE International Conference on Requirements Engineering*, Paris, France, 29 August – 2 September, Paris: IEEE.

Buxey, G. M., Slack, N. D. and Wild, R. (1973). Production Flow Line System Design – A Review. *AIIE Transactions*, 5: 37-48.

REFERENCES

Buzacott, J. A. (2013). The Design of Manufacturing Systems to Cope with Variability. In Handbook of Stochastic Models and Analysis of Manufacturing System Operations, New York: Springer, 1-28.

Buzacott, J. A. and Shantikumar, J. G. (1993). *Stochastic models of manufacturing systems*. Engelwood Coffs, New Jersey: Prentice Hall.

Capozzi, F. and Sacco, M. (2013). Exploiting Augmented Reality in Lean production: Opportunities and Challenges. *Advances in Sustainable and Competitive Manufacturing Systems: 23rd International Conference on Flexible Automation & Intelligent Manufacturing*, Porto, Portugal, 26-28 June, Switzerland: Springer International Publishing, pp. 1361-1373.

Carrascosa, M. (1995). *Variance of the output in a deterministic two-machine line*. MSc diss., Cambridge: Massachusetts Institute of Technology.

Caster Concepts (2012). *Go Beyond Standard with Mass Customization*. Michigan: Caster Concepts.

Chandra, C. and Kamrani, K. (2004). *Mass Customization: A Supply Chain Approach*. New York: Kluwer Academic/ Plenum Publishers.

Chituc, C. M. and Restivo, F. J. (2009). Challenges and Trends in Distributed Manufacturing Systems: Are wise engineering systems the ultimate answer?. *Second International Symposium on Engineering Systems*, Cambridge, Massachusetts, 15-17 June.

Chryssolouris, G. (2006). *Manufacturing Systems: Theory and Practice*. New York: Springer Science & Business Media.

Cicirello, V. and Smith, S. (2001). Ant colony control for autonomous decentralized shop floor routing. *Proceedings of the ISADS-2001 Fifth International symposium on Autonomous Decentralized Systems*, Dallas: IEEE.

Coenen, F. (2011). Data mining: past, present and future. *The Knowledge Engineering Review*, 26(1): 25-29.

Dagli, C. H. (1994). *Artificial Neural Networks for Intelligent Manufacturing*. London: Chapman and Hall.

Dallery, Y., David, R. and Xie, X. L. (1989). Approximate analysis of transfer lines with unreliable machines and finite buffers. *IEEE Transactions on Automatic Control*, 34: 943-953.

Daniel, V. R. and Guide, R. J. (2000). Production planning and control for remanufacturing: industry practice and research needs. *Journal of Operations Management*, 18: 467-483.

Dashkovskiy, S., Mironchenko, A. and Naujok, L. (2011). Autonomous Cooperation and Control. *In Logistics: Contributions and Limitations - Theoretical and Practical Perspectives*, Berlin: Springer-Verlag, 27-43.

Date, A. and Kurata, K. (2008). A property of neural networks of associative memory models with replacing units. *Artificial Life and Robotics*, 12: 291-294.

Davis, S. (1989). From future perfect: Mass customizing. *Planning Review*, 17 (2): 16-21.

de Beer, C. (2008). *Study the dynamics of self-controlling processes in production logistics systems based on discrete event simulation models*. Bremen: The University of Bremen.

Dirgo, R. (2006). *Look Forward Beyond Lean and Six Sigma*. Florida: J. Ross Publishing.

Djassemi, M. (2005). A Simulation Analysis of Factors Influencing the Flexibility of Cellular Manufacturing. *International Journal of Production Research*, 43 (10): 2101-2112.

Enginarlar, E., Li, J. and Meerkov, S. M. (2006). Lean Buffering in Serial Production Lines with Non-exponential Machines. *In Stochastic Modelling of Manufacturing Systems*, Berlin: Springer: 29-53.

Etman, L. F. and Rooda, J. E. (2000). *Process Time Variability and Distributions*. Eindhoven: Eindhoven University of Technology.

Fu, T. C., Cheung, T., Chung, K. F. and Ng, C. (2006). An Innovative Use of Historical Data for Neural Network Based Stock Prediction. *Proceedings of the 2006 Joint Conference on Information Sciences (JCIS 2006)*, Kaohsiung, Taiwan, ROC, 8-11 October.

Gershwin, S. B. (1993). Variance of the output of a tandem production system, queueing networks with finite capacity. *Second International Conference on Queueing Networks with Finite Capacity*. Amsterdam: Elsevier.

Gershwin, S. B. (1994). *Manufacturing Systems Engineering*. New Jersey: Prentice-Hall.

Gratton, C. and Jones, I. (2004). *Research Methods for Sport Studies*. London: Routledge.

Gronau, N., Theuer, H. and Lass, S. (2012). Evaluation of Production Processes Using Hybrid Simulation. *In Robust Manufacturing Control*, Berlin: Springer, 401-413.

Groover, M. P. (2010). *Fundamentals of Modern Manufacturing: Materials, Processes, and Systems*. 4th ed., New York: John Wiley & Sons.

Grundstein, S., Schukraft, S., Scholz-Reiter, B. and Freitag, M. (2015). Coupling order release methods with autonomous control methods – an assessment of potentials by literature review and discrete event simulation. *International Journal of Production Management and Engineering,* 3(1): 43-56.

Haik, B. and Aomar R. (2006). *Simulation-Based Lean Six-Sigma and Design for Six-Sigma*. New Jersey: John Wiley & Sons.

He, N., Zhang, D. Z. and Li, Q. (2014). Agent-based hierarchical production planning and scheduling in make-to-order manufacturing system. *International Journal of Production Economics*, 149: 117-130.

He, X. F., Wu, S. and Li, Q. L. (2007). Production variability of flow lines. *International Journal of Production Economics*, 107: 78-87.

Heilala, J. (1999). Use of simulation in manufacturing and logistics systems planning. *Proceedings of the VTT Manufacturing Technology*, Finland, 1-24.

Hillier, F. S. and Boling, R. W. (1967). Finite queues in series with exponential or Erlang service times – a numerical approach. *Operations Research*, 15: 286-303.

Hinich, M. J. and Talwar, P. P. (1975). A simple method for robust regression. *Journal of the American Statistical Association,* 70: 113–119.

Hitt, M. A, Keats, B. W. and DeMarie, S. M. (1998). Navigating in the new competitive landscape: Building strategic flexibility and competitive advantage in the 21st century. *Academy of Management Executive*, 12(4): 22-42.

Hoad, K., Robinson, S. and Davies, R. (2008). Automating warm-up length estimation. *Proceedings of the 2008 Winter Simulation Conference*. New York: IEEE.

Holland, P. W. and Welsch, R. E. (1977). Robust regression using iteratively reweighted least-squares. *Communications in Statistics - Theory and Methods,* A6: 813–827.

Hopp, W. (2008). *Supply Chain Science*. Illinois: Waveland Press.

Hopp, W. and Spearman, M. (2011). *Factory Physics.* 3rd ed., Illinois: Waveland Press.

Huber, P. J. (1981). *Robust Statistics*. Ney York: John Wiley & Sons.

Hunt, G. C. (1956). Sequential arrays of waiting lines. *Operations Research*, 4: 674-683.

Industrial Technology Centre (2004). *Principles of Lean Thinking: Tools & Techniques for Advanced Manufacturing*. Ottawa: National Research Council Canada.

Jacobs, D. and Meerkov, S. M. (1995). A system-theoretic property of serial flow lines: improvability. *International Journal of Systems Science*, 26 (4): 755-785.

Jacobs, J. H., Etman, L. F., van Campen, E. J. and Rooda, J. E. (2003). Characterization of Operational Time Variability Using Effective Process Times. *IEEE Transactions on Semiconductor Manufacturing*, 16(3): 511-520.

Jaikumar, R. (1993). 200 years to CIM. *IEEE Spectrum*, 30(9): 26-27.

Jain, A., Jain, P. K., Chan, F. T. and Singh S. (2013). A review on manufacturing flexibility. *International Journal of Production Research*, 51(19): 5946-5970.

Jasti, N. V. and Kodali, R. (2015). Lean production: literature review and trends. *International Journal of Production Research*, 53(3): 867-885.

Kang, N., Zheng, L. and Li, J. (2015). Analysis of multi-product manufacturing systems with arbitrary processing times. *International Journal of Production Research*, 53(3): 983-1001.

Kang, P. S., Duffy, A., Shires, N., Smith, T. and Novels, M. (2014). Lean Cables – A Step towards Competitive, Sustainable and Profitable Processes. *Proceedings of the 63rd IWCS Conference*, RI, USA, 09-12 Nov, 310 – 317.

Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms*. New Jersey: Wiley.

Kennedy, P. (2009). *How to combine multiple research methods: Practical Triangulation*. Johnny Holland [Online], 20 August. Available from: <http://johnnyholland.org/ > [Accessed 02 October 2013].

Khalil, R. (2005). *Predicting the effect of variability on the efficiency of flow processing system*s. PhD diss., Leicester: De Montfort University.

King, G., Keohane, O. and Verba, S. (1994). *Designing Social Inquiry. Scientific Inference in Qualitative Researc*h. New Jersey: Princeton University Press.

Knott, A. (1970). The Inefficiency of a series of Work-stations – A Simple Formula. *International Journal of Production Research*, 8 (2): 109-119.

Kock, A. A., Etman, L. F. and Rooda, J. E. (2008). Effective process times for multi-server flowlines with finite buffers. *IIE Transactions*, 40: 177-186.

Kuma, J. W. (1984). *Basic Statistics for the Health Science*s. Palo Alto: Mayfield Publishing Co.

Lagershausen, S. and Tan, B. (2015). On the Exact Inter-departure, Inter-start, and Cycle Time Distribution of Closed Queueing Networks Subject to Blocking. *IIE Transactions*, 47(7): 673-692.

Laguna, M. and Marklund, J. (2013). *Business Process Modelling, Simulation and Design*. Boca Raton: CRC Press.

Law, A. M. (2007). *Simulation modelling and analysis*. 4th ed. New York: McGraw-Hill.

Law, A M. and McComas, M. (1990). Secrets of successful simulation studies. *Industrial Engineering*, 22(5): 47-72.

Leu, Y., Matheson, L. A. and Rees, L. P. (1996). Sequencing Mixed-model Assembly Lines with Genetic Algorithms. *Computers & Industrial Engineering*, 30(4): 1027-1036.

Li, J., Blumenfeld, D., Huang, N. and Alden, M. (2009). Throughput analysis of production systems: recent advances and future topics. *International Journal of Production Research*, 47(14): 3823-3851.

Li, J. and Meerkov, S. M. (2003). Due-Time Performance of Production Systems with Markovian Machines. *In Analysis and Modelling of Manufacturing Systems*, Dordrecht: Kluwer Academic, 221-253.

Li, J. and Meerkov, S. M. (2005). Evaluation of Throughput in Serial Production Lines with Non-exponential Machines. *In Analysis, Control and Optimization of Complex Dynamic Systems*, New York: Springer, 61-90.

Li, J. and Meerkov, S. M. (2009). *Production systems engineering*. New York: Springer.

Li, J., Meerkov, S. M. and Zhang, L. (2013). Production systems engineering: Review and recent developments. *In Handbook of Stochastic Models and Analysis of Manufacturing System Operations*, New York: Springer, 167-210.

Liker, J. (2004). *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*. New York: McGraw-Hill.

Martin, W. E. and Bridgmon, K. D. (2012). *Quantitative and Statistical Research Methods: From Hypothesis to Results*. San Francisco: John Wiley & Sons.

Medhi, J. (1991*). Stochastic Models in Queuing Theory*. Boston: Academic Press.

Meerkov, S. M. and Yan, C. (2014). Production Lead Time in Serial Lines: Evaluation, Analysis, and Control. *IEEE Transactions on Automation Science and Engineering*, 99: 1-13.

Mehrabi, M. G., Ulsoy, A. G. and Koren, Y. (2000). Reconfigurable manufacturing systems: Key to future manufacturing. *Journal of Intelligent manufacturing*, 11(4): 403-419.

Mendenhall, W., Beaver, R. and Beaver, B. (2012). *Introduction to Probability and Statistics*. Boston: Cengage Learning.

Meystel, A and Albus, J. (2001). *Intelligent Systems: Architecture, Design and Control*. New York: John Wiley and Sons.

Miltenburg, G. J. (1987). Variance of the number of units produced on a transfer line with buffer inventories during a period of length t. *Naval Research Logistics*. 34: 811-822.

Moore, H. R. II. (1977). *Robust Regression using Maximum-Likelihood Weighting and Assuming Cauchy-Distributed Random Error*. MSc diss., Monterey: Naval Postgraduate School.

Mukhopadhyay, S. K. (2015). *Production planning and control: text and cases*. 3rd ed., Delhi: PHI Learning Private Limited.

Muth, E.J. (1973). The Production Rate of a Series of Work Stations with Variable Service Times. *International Journal of Production Research*, 11: 155-169.

Muth, E. J. (1987). An Update on Analytical Models of Serial Transfer Lines. *Research Report*, 87-15.

Nath, R. (2009). *Philosophy of Artificial Intelligence: A Critique of the Mechanistic Theory of Mind*. Florida: Universal Publishers.

Nehmzow, U. (2009). *Robot Behaviour: Design, Description, Analysis and Modelling*. London: Springer-Verlag.

Norris, J. R. (1997). *Markov Chains*. Cambridge: Cambridge University Press.

Oates, T. and Jensen, D. (1998). Large datasets lead to overly complex models: an explanation and a solution. *In Proceedings of Fourth International Conference on Knowledge Discovery and data Mining*. California: AAAI Press.

OECD. (2007). *Understanding the Brain: The Birth of a Learning Science*. Paris: OECD.

Oztemel, E. (2010). *Intelligent Manufacturing Systems in Artificial Intelligence Techniques for Networked Manufacturing Enterprises Management*. London: Springer-Verlag, 1-41.

Pandey, P. (2008). *Computer Aided Process Planning*. Lecture Notes, Delhi: Indian Institute of Technology.

Panneerselvam, R. (2012). *Production and operations management*. 3rd ed., Delhi: PHI Learning Private Limited.

Paolucci, M. and Sacile, R. (2004). *Agent-Based Manufacturing and Control Systems: New Agile Manufacturing Solutions for Achieving Peak Performance*. Florida: CRC Press.

Papadopoulos, C. T., O'Kelly, M. E., Vidalis, M. J. and Spinellis, D. (2009). *Analysis and Design of Discrete Part Production Lines*. New York: Springer.

Papadopoulos, H. T. (1996). An analytic formula for the mean throughput of K-station production lines with no intermediate buffers. *European Journal of Operational Research*, 91: 481-494.

Patel, J. K. and Read, C. B. (1996). *Handbook of the Normal Distribution*. Florida: CRC Press.

Pine, B. J. II, (1993). *Mass Customisation. The New Frontier in Business*, Boston: Harvard Business School Press.

Punch, K. (2000). *Developing Effective Research Proposals*. London: Thousand Oaks.

Quadt, D. and Kuhn, H. (2005). Conceptual framework for lot-sizing and scheduling of flexible flow lines. *International Journal of Production Research*, 43(11): 2291-2308.

Quadt, D. and Kuhn, H. (2007). A taxonomy of flexible flow line scheduling procedures. *European Journal of Operational Research*, 178(3): 686-698.

Rajkumar, T. and Bardina, J. (2003). Training data requirement for a neural network to predict aerodynamic coefficients, Independent Component Analyses, Wavelets, and Neural Networks. *Proceedings of the SPIE*, 5102: 92-103.

Raol, J. R. and Gopal, A. K. (2013). *Mobile Intelligent Autonomous Systems*. Florida: CRC Press.

Rekersbrink, H. (2012). *Methods for self-controlling routing of autonomous logistic objects*. PhD diss., Germany: Universität Bremen.

Rey, W. J. J. (1983). *Introduction to Robust and Quasi-robust Statistical Methods*. Berlin: Springer-Verlag.

Rippel, D., Harjes, F. and Scholz-Reiter, B. (2010). Modelling a Neural Network Based Control for Autonomous Production Systems. *In Artificial Intelligence and Logistics (AILog) Workshop at the 19th European Conference on Artificial Intelligenc*e, Amsterdam, Netherlands, 49-54.

Robinson, S. (2004). *Simulation; the Practice of Model Development and Use*. New York: John Wiley & Sons Ltd.

Sankar, N. P., Babu, N. R. and Chetty, O. V. (1997). Scheduling of Flexible Flow Lines with Buffer and Handling System Constraints. *Proceedings of 17th All India Manufacturing Technology: Design and Research Conference*, New Delhi: Allied Publishers.

Saunders, M., Lewis, P. and Thornhill, A. (2009). *Research methods for business students*, 5th ed., Essex: Pearson Education Limited.

Scallan, P. (2003). *Process Planning*. Oxford: Butterworth-Heinemann.

Scherer, E. (1998). The Reality of Shop Floor Control – Approaches to Systems Innovation. *In Shop Floor Control – A Systems Perspective*. Berlin: Springer Verlag.

Scholz-Reiter, B., de Beer, C. D., Freitag, M. and Jagalski, T. (2008a). Bio-inspired and pheromone-based shop-floor control. *International Journal of Computer Integrated Manufacturing*, 21(2): 201-205.

Scholz-Reiter, B. and Freitag, M. (2007). Autonomous processes in assembly systems. CIRP Annals, 56(2): 712-729.

Scholz-Reiter, B., Freitag, M., Beer, C. and Jagalski, T. (2005). Modelling and Analysis of Autonomous Shop Floor Control. *Proceedings of the 38th CIRP International Seminar on Manufacturing Systems*, Florianopolis, Brazil, 16-18 May.

Scholz-Reiter, B., Freitag, M. de Beer, C. and Jagalski, T. (2006). The Influence of Production Networks' Complexity on the Performance of Autonomous Control Methods. *In Proceedings of the 5th CIRP International Seminar on Computation in Manufacturing Engineering,* Ischia, Italy, 25-28 July.

Scholz-Reiter, B., Görges, M., Jagalski, T. and Mehrsai, A. (2009a). Autonomously controlled production systems—Influence of autonomous control level on logistic performance, *CIRP Annals - Manufacturing Technology*, 58(1): 395 - 398.

Scholz-Reiter, B., Görges, M., Jagalski, T. and Mehrsai, A. (2009b). Modelling and analysis of autonomously controlled production networks. *Proceedings of the 13th IFAC Symposium on Information Control Problems in Manufacturing*, Moscow, Russia, 850-855.

Scholz-Reiter, B., Görges, M., Jagalski, T. and Naujok, L. (2010). Modelling and analysis of an autonomous control method based on bacterial chemotaxis. *Proceedings of the 43rd CIRP International Conference on Manufacturing Systems*, Wien: Neuer Wissenschaftlicher Verlag, 699-706.

Scholz-Reiter, B., Jagalski, T. and Bendul, J. (2008b). Autonomous control of a shop floor based on bee's foraging behaviour. *Dynamics in Logistics*. Berlin: Springer, 415-423.

Scholz-Reiter, B., Rekersbrink, H. and Görges, M. (2010). Dynamic Flexible Flow Shop Problems – Scheduling Heuristics vs. Autonomous Control. *CIRP Annals - Manufacturing Technology*, 59: 465-468.

Sennott, L., Van Oyen, M. and Iravani, M. (2004). Optimal Dynamic Assignment of a Flexible Worker on an Open Flow line with Specialists. *Operations research, Management science*, 47(1): 69-72.

Sethi, A. K. and Sethi, S. P. (1990). Flexibility in Manufacturing: a survey. *International Journal of Flexible Manufacturing Systems*, 2: 289-328.

Shankar, R. (2009). *Process Improvement Using Six Sigma: A DMAIC Guide*. Milwaukee: ASQ Quality Press.

Shao, J and Tu, D. (2006). *The Jackknife and Bootstrap (Springer Series in Statistics)*. London: Springer-Verlag.

Sheth-Voss, P. A., Willemain, T. R. and Haddock, J. (2005). Estimating the steady state mean from short transient simulations. *European Journal of Operational Research*, 162(2): 403-417.

Shipp, S. S., Gupta, N., Lal, B., Scott, J. A., Weber, C. L., Finnin, M. S., Blake, M., Newsome, S. and Thomas S. (2012). *Emerging Global Trends in Advanced Manufacturing*. Virginia: Institute for Defense Analyses.

Shortle, J., Rebennack, S. and Glover, F. W. (2014). Transmission-Capacity Expansion for Minimizing Blackout Probabilities. *IEEE Transactions on Power Systems*, 29(1): 43-52.

Silveira, G., Borenstein, D. and Fogliatto, F. (2001). Mass customization: Literature review and research directions. *International Journal of Production Economics*, 72: 1-13.

Singh, N. (1996). *Systems Approach to Computer-Integrated Design and Manufacturing*. New York: John Wiley & Sons.

Slack, N. D. and Wild, R. (1980). *The Nature, Performance and Operating Characteristics of Series Production Systems*. *International Journal of Operations & Production Management*, 1(2): 95 – 107.

Soman, C., Donk, D. and Gaalman, G. (2004). Combined make-to-order and make-to-stock in a food production system. *International Journal of Production Economics*, 90(2): 223-235.

Spinellis, D. D. and Papadopoulos, C. T. (2000). Stochastic Algorithms for Buffer Allocation in Reliable Production Lines. *Mathematical Problems in Engineering*, 5: 441-458.

Stelson, K. A., Liang, S. Y., Oba, F. and Arai, T. (1996). *Proceedings of the Japan-U.S.A. Symposium on Flexible Automation*. New York: American Society of Mechanical Engineers.

Tan, B., Gershwin, S. B., Liberopoulos, G., Meerkov, S. M. and Papadopoulos, C. T. (2015). Advances in stochastic models of manufacturing and service operations. *Annals of Operations Research*, 231(1).

Tewksbury, R. (2009). Qualitative versus Quantitative Methods: Understanding Why Qualitative Methods are Superior for Criminology and Criminal Justice. *Journal of Theoretical and Philosophical Criminology*, 1: 38-58.

Theuer, H., Gronau, N. and Lass, S. (2013). The Impact of Autonomy on Lean Manufacturing Systems. *In Advances in Sustainable and Competitive Manufacturing Systems*, Cham: Springer, 1413-1423.

Thurmond, V. (2001). The Point of Triangulation. *Journal of Nursing Scholarship*, 33(3), 254–256.

Tsutsui, S. and Liu, L. (2007). Cunning ant system for quadratic assignment problem with local search and parallelization. *In Pattern Recognition and Machine Intelligence*, Berlin, Germany: Springer, 269-278.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Boston: Addison-Wesley Publishers.

Ulrich, H. and Probst, G. (1988). *Instructions for holistic thinking and acting*. Bern-Stuttgart-Wien: Haupt Verlag AG.

Upton, D. M. (1994). The management of manufacturing flexibility. *California Management Review*, 36(2): 72-89.

Van Brackel, T. (2009). *Adaptive Control Flexible Manufacturing Systems Workshop*. Wiesbaden: Gabler Verlag.

Wang, J., Zhong, X. and Li, J. (2014). Modelling and Analysis of Care Delivery Services within Patient Rooms: A System-Theoretic Approach. IEEE Transactions on Automation Science and Engineering, 11(2): 379-393.

Weiss, S. M. (1998). *Predictive Data Mining: A Practical Guide*. San Francisco: Morgan Kaufmann Publishers.

Wiendahl, H. P., Elmaraghy, H. A., Nyhuis, P., Zah, M. F., Wiendahl, H. H., Duffie, N. and Brieke, M. (2007). Changeable Manufacturing – Classification Design and Operation. *CIRP Annals – Manufacturing Technology*, 56(2): 783-809.

Windt, K. and Becker, T. (2009). Applying Autonomous Control Methods in Different Logistic Processes - A comparison by Using an Autonomous Control Application Matrix. *Proceedings of the 17th Mediterranean Conference on Control & Automation*, Thessaloniki, Greece, 24-26 June, 1451-1415.

Windt, K., Becker, T. and Jeken, O. (2010). A classification pattern for autonomous control methods in logistics. *Logistics Research,* 2(2): 109-120.

Windt, K. and Jeken, O. (2009). Allocation Flexibility - A New Flexibility Type as an Enabler for Autonomous Control in Production Logistics. *42nd CIRP Conference on Manufacturing Systems*, Grenoble, France, 3-5 June.

Windt, K. and Jeken, O. (2010). Autonomous Product Manufacturing Cycle – An Integrated Approach to Process Planning and Production Control. *2010 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Macao, China, 7-10 December, 1286-1290.

Windt, K., Philipp, T. and Bose, F. (2008). Complexity cube for the characterisation of complex production systems. *International Journal of Computer Integrated Manufacturing*, 21: 195-200.

Womack, J. P., Jones, D. T. and Roos, D. (1991). *The Machine That Changed the World: The Story of Lean Production*. New York: HarperCollins.

Yaloaoui, A., Chehade, H., Yalaoui, F. and Amodeo, L. (2012). *Optimization of Logistics*. Wiley: New Jersey.

Yang, Q. and Wu, X. (2006). 10 challenging problems in data mining research. *International Journal of Information Technology & Decision Making*, 5(4): 597-604.

Zeng, J., Lin, I-J., Hoarau, E. and Dispoto, G. (2009). Numerical Simulation and Analysis of Commercial Print Production Systems. *25th International Conference on Digital Printing Technologies (NIP25)*, Kentucky, United States, 20 September.

Zhang, N. (2009). On Assortment and Implementation of Modern Mass Customization of Apparel. *Asian Social Media*, 5(2).

Zhou, Z., Wang, H. and Lou, P. (2010). *Manufacturing Intelligence for Industrial Engineering*. New York: Engineering Science Reference.

Zozom, A., Hodgson, T., King, R., Weintraub, A. and Cormier, D. (2003). Integrated job release and shop-floor scheduling to minimize WIP and meet due-dates. *International Journal of Production Research*, 41(1): 31–45.

BIBLIOGRAPHY

Ahlemeyer-Stubbe, A. and Coleman, S. (2014). *A Practical Guide to Data Mining for Business and Industry*. New York: John Wiley & Sons.

Anderson, D. R., Sweeney, D. J., Williams, T. A. Freeman, J. and Shoesmith, E. (2010). *Statistics for Business and Economics*. 2nd ed., Hampshire: CENGAGE Learning Business Press.

Bolch, G. and Greiner, S. (1998). *Queueing Networks and Markov Chains: Modelling and Performance Evaluation with Computer Science Applications*. New York: John Wiley & Sons.

Dasgupta, A. Sun, Y. V., König, I. R., Bailey-Wilson, J. E. and Malley, J. D. (2011). Brief Review of Regression-Based and Machine Learning Methods in Genetic Epidemiology: The Genetic Analysis Workshop 17 Experience. *Genetic Epidemiology*, 35(1): S5–11.

Esener, E., Firat, M. and Gurler, B. (2013). Predictive Modelling of Drawbeads Using Regression and Neural Network Techniques. *7th International Conference and Exhibition on Design and Production of Machines and Dies/Molds*, Antalya, Turkey.

Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques, *Informatica*, 31, 249-268.

Lee, W. (1999). A Data Mining Framework for building intrusion detection models, *Proceedings of the 1999 IEEE Symposium on Security and Privacy*. 09-12 May, Oakland, CA, 120-132.

Manyonge, L.M. (2012). *Autonomous Finite Capacity Scheduling using Biological Control Principles*. PhD diss., Leicester: De Montfort University.

Mitchell, H. B. (2012). *Data Fusion: Concepts and Ideas*. 2nd ed., Berlin: Springer.

Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. 2nd ed., New York: Cambridge University Press.

Ripley, B. D. (2002). *Statistical Data Mining*. New York: Springer.

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58(1): 276-288.

APPENDICES

APPENDIX A – Scope of the Research in Evaluative Modelling

Refer to Section 1.5.1

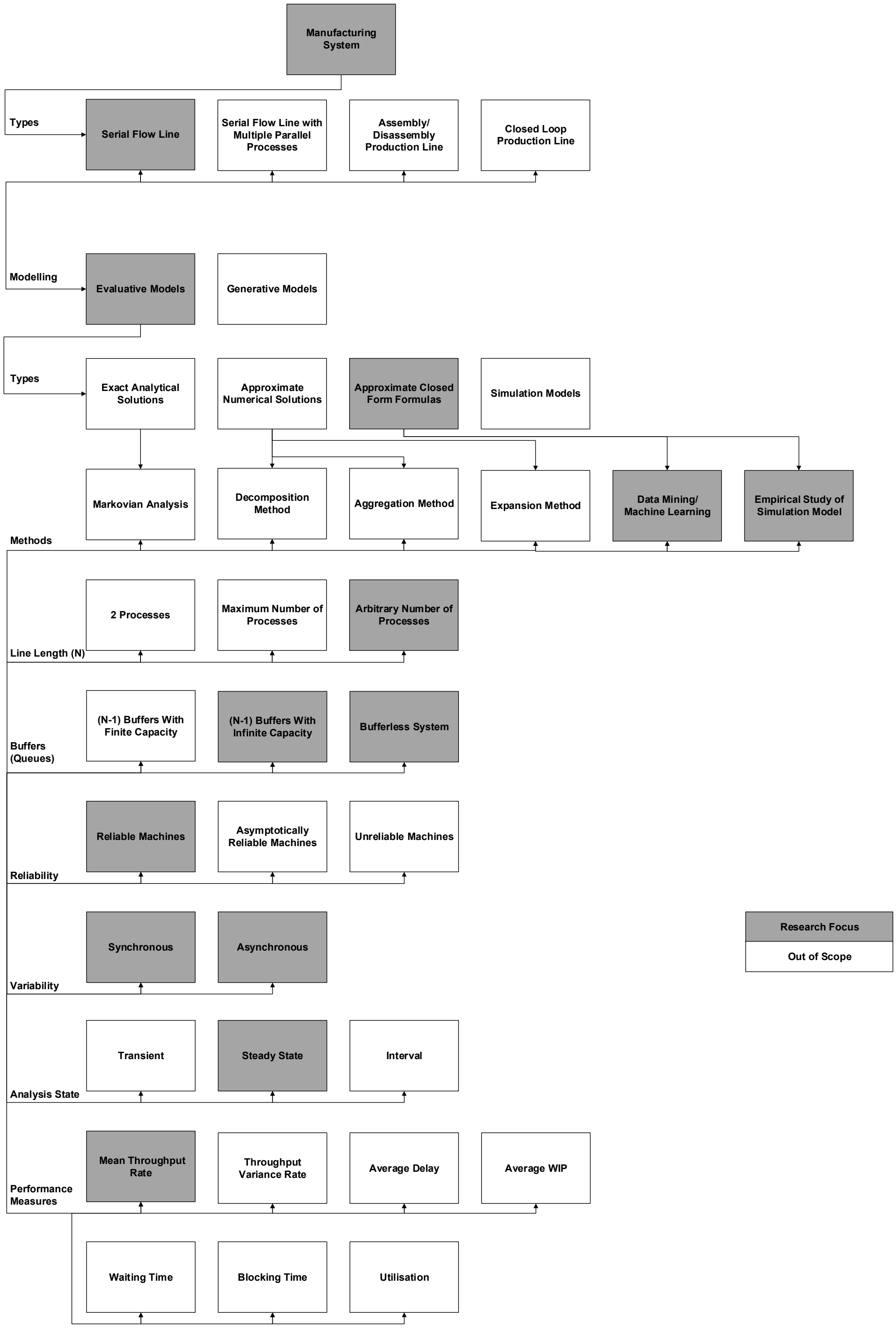


Figure A.1: Scope of the Research in Evaluative Modelling

APPENDIX B – Synthetic Data Set $I / II - S - 1$ for Asynchronous Flow Line

Refer to Section 4.5.2.1.1

Table B.1: Data Set $I/II-S-1$

Set	Length	Mean Processing Time	Coefficient of Variation			
1	2	1	0.01	Single Factor	*Processing Time*	*Included*
1	2	2	0.01		Coefficient of Variation	Not Included
1	2	3	0.01		Length	Not Included
1	2	4	0.01			
1	2	5	0.01			
1	2	6	0.01			
1	2	7	0.01			
1	2	8	0.01			
1	2	9	0.01			
1	2	10	0.01			
1	2	1	0		*Processing Time*	*Included*
1	2	2	0		Coefficient of Variation	Not Included
1	2	3	0		Length	Not Included
1	2	4	0			
1	2	5	0			
1	2	6	0			
1	2	7	0			
1	2	8	0			
1	2	9	0			
1	2	10	0			
1	2	1	0.025		Processing Time	Not Included
1	2	1	0.05		*Coefficient of Variation*	*Included*
1	2	1	0.075		Length	Not Included
1	2	1	0.1			
1	2	1	0.25			
1	2	1	0.5			
1	2	1	0.75			
1	2	1	1			
1	3	1	0.01		Processing Time	Not Included
1	4	1	0.01		Coefficient of Variation	Not Included
1	5	1	0.01		*Length*	*Included*
1	6	1	0.01			
1	7	1	0.01			
1	8	1	0.01			
1	9	1	0.01			
1	10	1	0.01			
1	11	1	0.01			
1	12	1	0.01			
1	13	1	0.01			
1	14	1	0.01			
1	15	1	0.01			
1	16	1	0.01			
1	17	1	0.01			
1	18	1	0.01			
1	19	1	0.01			
1	20	1	0.01			
1	21	1	0.01			
1	3	1	0		Processing Time	Not Included
1	4	1	0		Coefficient of Variation	Not Included
1	5	1	0		*Length*	*Included*
1	6	1	0			
1	7	1	0			
1	8	1	0			
1	9	1	0			
1	10	1	0			
1	11	1	0			
1	12	1	0			
1	13	1	0			
1	14	1	0			
1	15	1	0			
1	16	1	0			
1	17	1	0			
1	18	1	0			
1	19	1	0			
1	20	1	0			
1	21	1	0			

APPENDIX C – Data Cleaning – Smoothing Results

Refer to Section 4.5.2.3

Smoothing of the data output, i.e., simulated TR performed poorly with all methods. Local and robust regression models of weighted linear least squares of the 2nd degree polynomial and Savitzky-Golay filter performed nearly the same and they gave better accuracy than other models for a flow line model of one process only. This has not been the case when the flow line length increased to 2 processes, and all methods rounded up to fairly the same errors. Table C.1 gives the Minimum, Average and Maximum Absolute Percentage Errors (MINAPE), (MAPE) and (MAXAPE), respectively, between the smoothed and simulated throughput rate for the variability scenarios in data set $I/II-S-1$ and $I/II-S-2$.

Table C.1: Smoothing Errors for Data Set $I/II-S-1$ and $I/II-S-2$

			$I/II-S-1$			
	Moving Average	Local Regres. 1st_D	Robust Regres. 1st_D	Local Regres. 2nd_D	Robust Regres. 2nd_D	Savitzky-Golay filter
MINAPE	0%	6.38%	0%	0%	0%	0%
MAPE	502%	928%	654%	458%	459%	458%
MAXAPE	1046%	2714%	1534%	984%	984%	984%
			$I/II-S-2$			
	Moving Average	Local Regres. 1st_D	Robust Regres. 1st_D	Local Regres. 2nd_D	Robust Regres. 2nd_D	Savitzky-Golay filter
MINAPE	60.20%	51%	78%	99%	99%	99%
MAPE	830%	829%	830%	830%	830%	830%
MAXAPE	1826%	2657%	2315%	1823%	1823%	1823%
			Overall			
	Moving Average	Local Regres. 1st_D	Robust Regres. 1st_D	Local Regres. 2nd_D	Robust Regres. 2nd_D	Savitzky-Golay filter
MINAPE	0%	6.38%	0%	0%	0%	0%
MAPE	830%	829%	829%	830%	830%	830%
MAXAPE	1826%	2714%	2315%	1823%	1823%	1823%

APPENDIX D – Phase III Model Building Flowchart

Refer to Section 4.5.3.2.1

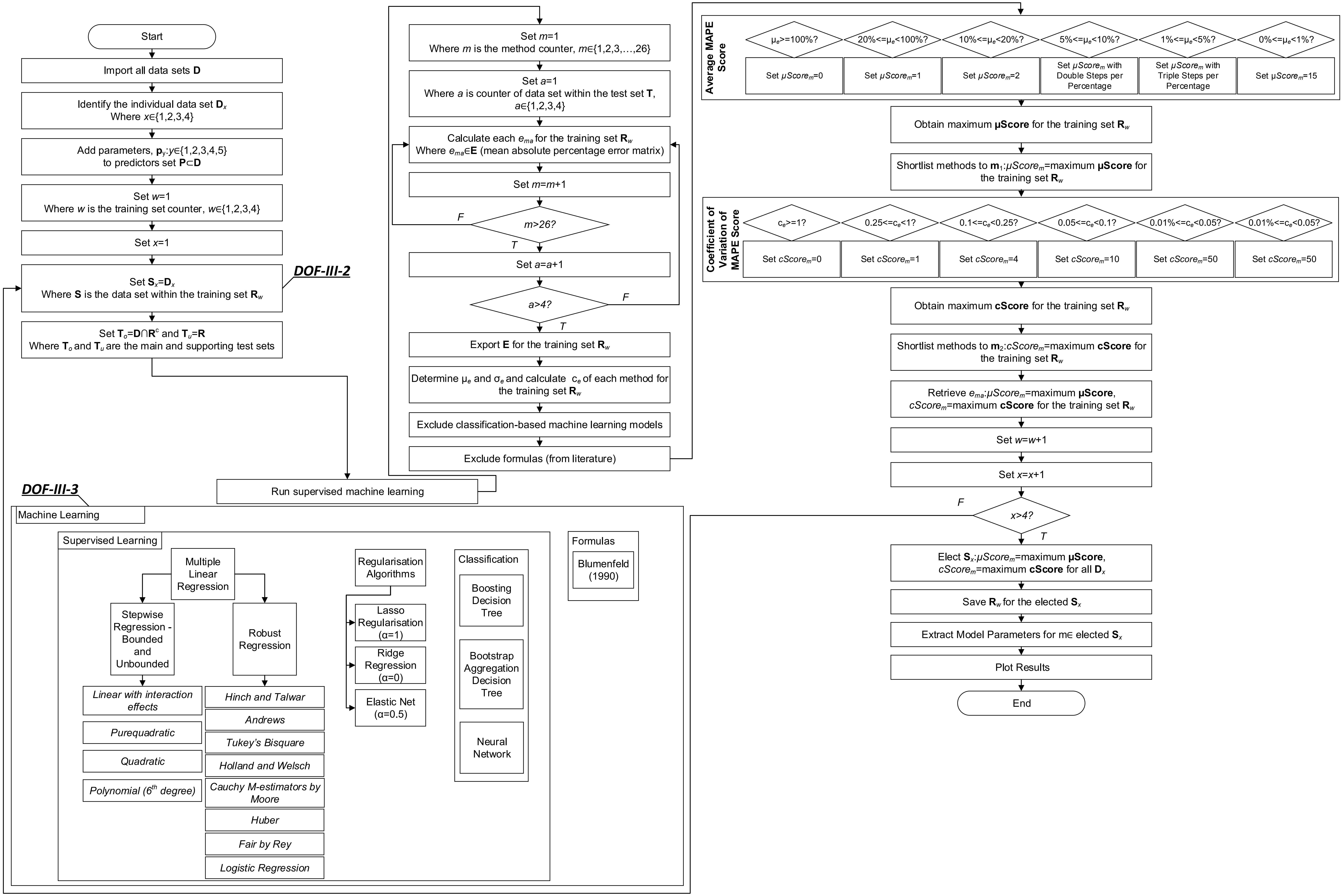


Figure D.1: Model Building Flowchart – Synchronous Flow Line

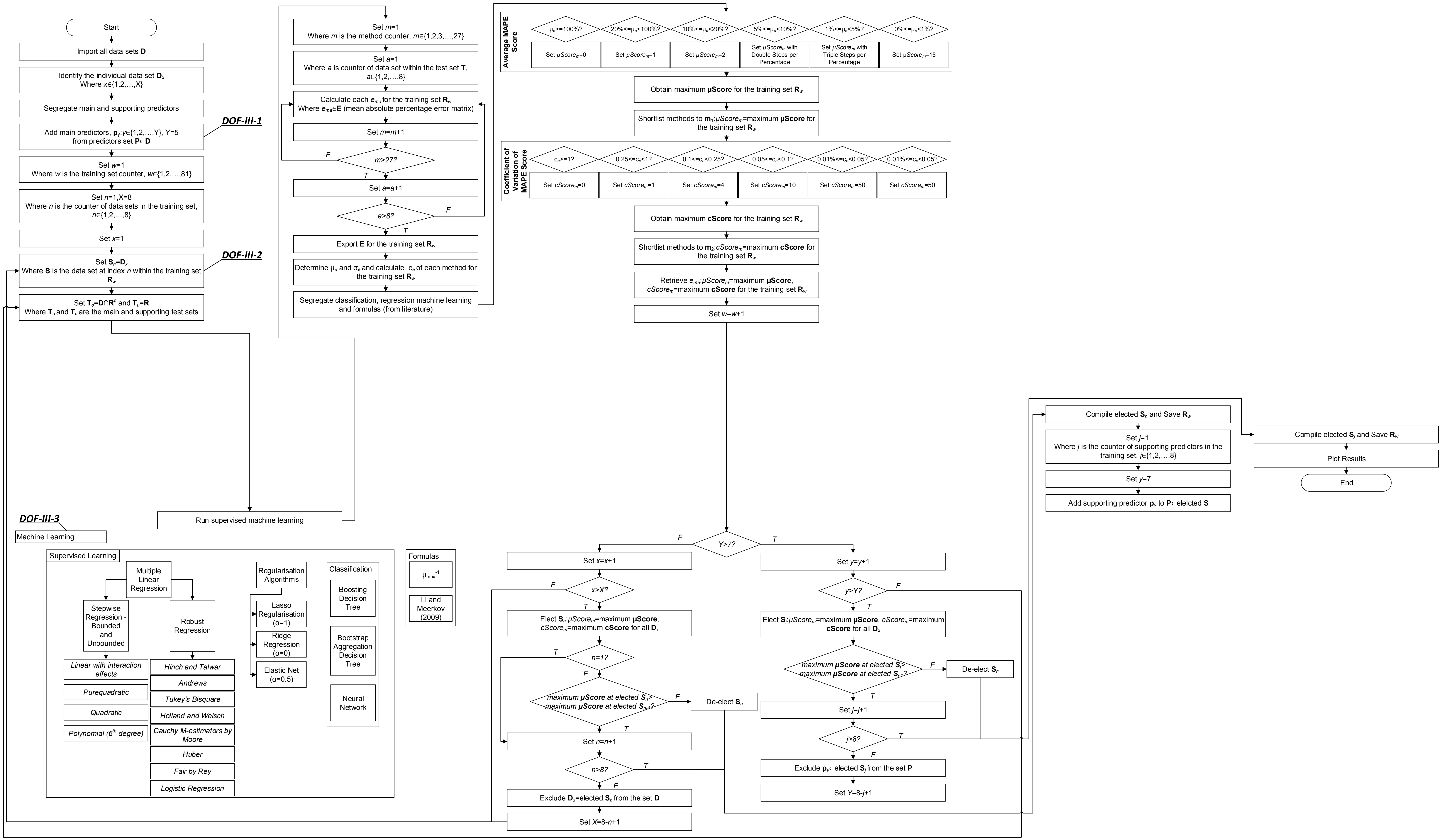


Figure D.2: Model Building Flowchart – Asynchronous Flow Line

APPENDIX E – Phase II – Correlation and ANOVA Analysis of the Relationship between Minimum, Average and Coefficient of Variation of Mean Processing Time Terms and the Throughput Rate

Refer to Section 6.3.5

Table E.1: Correlation Analysis of the Relationship between Minimum, Average and Coefficient of Variation of Mean Processing Time Terms and the Throughput Rate

Parameter Term	TR_1	TR_2	TR_3	TR_4
μ	-0.99	-0.75	-0.76	-0.93
μ^{-1}	0.72	0.65	0.66	0.92
$\log \mu$	-0.90	-0.71	-0.72	-0.92
$\log \frac{1}{\mu}$	0.90	0.71	0.72	0.92
e^{μ}	-0.37	-0.85	-0.85	-1.00
$e^{\mu^{-1}}$	0.71	0.65	0.66	0.92
$\mu_{\min}$	-0.99	-0.75	-0.76	-0.93
$\mu_{\min}^{-1}$	0.46	0.26	0.27	0.38
$\log \mu_{\min}$	-0.84	-0.55	-0.56	-0.73
$\log \frac{1}{\mu_{\min}}$	0.84	0.55	0.56	0.73
$e^{\mu_{\min}}$	-0.36	-0.71	-0.72	-0.50
$e^{\mu_{\min}^{-1}}$	0.38	0.21	0.22	0.32
c	0.72	0.65	0.66	0.92
c^{-1}	-0.59	-0.90	-0.90	-0.74
$\log c$	**0.97**	**0.93**	**0.94**	**0.99**
$\log \frac{1}{c}$	**-0.97**	**-0.93**	**-0.94**	**-0.99**
e^{c}	0.46	0.58	0.59	0.91
$e^{c^{-1}}$	-0.24	-0.53	-0.54	-0.35

Table E.2: ANOVA Analysis of the Relationship between Minimum, Average and Coefficient of Variation of Mean Processing Time Terms and the Throughput Rate

Experiment 1					
Source	*DF*	*Adj SS*	*Adj MS*	*f-value*	*p-value*
Regression	5	509104	101821	1.12E06	0
$\log \mu$	**1**	**212**	**212**	**2341.23**	**0**
$e^{\mu^{-1}}$	**1**	**5**	**5**	**51.97**	**0**
c	**1**	**6**	**6**	**69.59**	**0**
c^{-1}	**1**	**126**	**126**	**1388.46**	**0**
$\log c$	1	374	374	4127.49	0
Error	34	3	0		
Total	39	509107			
Experiment 2					
Source	*DF*	*Adj SS*	*Adj MS*	*f-value*	*p-value*
Regression	1	6.82754	6.82754	1.87E+03	0
c	**1**	**6.82754**	**6.82754**	**1868.9**	**0**
Error	23	0.08402	0.00365		
Total	24	6.91156			
Experiment 3					
Source	*DF*	*Adj SS*	*Adj MS*	*f-value*	*p-value*
Regression	4	7.34273	1.83568	4.12E+02	0
μ	**1**	**0.16199**	**1.62E-01**	**36.34**	**0**
$e^{\mu_{\min}^{-1}}$	**1**	**0.06923**	**6.92E-02**	**15.53**	**0.001**
Error	20	0.08914	0.00446		
Total	24	7.43188			
Experiment 4					
Source	*DF*	*Adj SS*	*Adj MS*	*f-value*	*p-value*
Regression	3	150.423	50.141	1.14E+04	0
$e^{\mu_{\min}^{-1}}$	1	0.013	0.013	2.96	0.094
c^{-1}	**1**	**0.712**	**0.7116**	**162.22**	**0**
$\log \frac{1}{c}$	**1**	**0.723**	**0.7233**	**164.88**	**0**
Error	37	0.162	0.0044		
Total	40	150.585			

APPENDIX G – Phase III – MAPE of the Individual Test Set of Data Mining Models for Synchronous Flow Lines

Refer to Section 6.4.1

Table G.1: MAPE of the Individual Test Set of Data Mining Models for Synchronous Flow Lines with the Training Set *I-II-S-1*

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_u, u=1$ *I-II-S-1*	Test Set $T_o, o=1$ *I-II-S-2*	Test Set $T_o, o=2$ *I-II-S-3*	Test Set $T_o, o=3$ *I-II-S-4*	μ_e	c_e
1	Decision Tree	Boosting			0.00%	23.47%	0.00%	138.45%	**40.48%**	**1.64**
2		Bootstrap Aggregating			23.42%	229.56%	573.69%	796.78%	**405.86%**	**0.85**
3	Neural Network	Feedforward			0.16%	179.76%	368.27%	240.18%	**197.09%**	**0.78**
4	Blumenfeld (1990)				0.37%	1.73%	3.73%	4.67%	**2.63%**	**0.74**
5	Multiple Linear Regression	Robust	Tukey's Bisquare		1.59%	28.88%	167.74%	220.44%	**104.66%**	**1.01**
6			Andrews		1.59%	28.89%	167.74%	220.44%	**104.67%**	**1.01**
7			Cauchy M-estimators by Moore		0.13%	23.27%	107.19%	145.98%	**69.14%**	**1.00**
8			Fair by Rey		0.09%	22.86%	106.18%	144.31%	**68.36%**	**1.00**
9			Huber		0.12%	23.11%	106.71%	145.22%	**68.79%**	**1.00**
10			Logistic Regression		0.12%	23.10%	106.70%	145.21%	**68.78%**	**1.00**
11			Hinch and Talwar		1.59%	28.86%	167.58%	220.27%	**104.58%**	**1.01**
12			Holland and Welsch		0.15%	23.29%	107.12%	145.91%	**69.12%**	**1.00**
13		Regularisation	Lasso		0.57%	22.74%	91.79%	125.06%	**60.04%**	**0.97**
14			Ridge Regression		0.09%	22.63%	105.81%	143.41%	**67.99%**	**1.00**
15			Elastic Nets		0.57%	22.74%	91.79%	125.06%	**60.04%**	**0.97**
16		Stepwise	Linear with Interactions of Linear Terms	Bounded Steps	0.05%	23.24%	108.75%	148.11%	**70.04%**	**1.00**
17				Unbounded Steps	0.05%	23.29%	108.99%	148.01%	**70.09%**	**1.00**
18			Linear and Squared Terms	Bounded Steps	0.02%	23.68%	110.92%	150.70%	**71.33%**	**1.00**
19				Unbounded Steps	0.02%	23.68%	110.92%	150.70%	**71.33%**	**1.00**
20			Linear and Squared Terms including Interactions	Bounded Steps	0.02%	23.68%	110.92%	150.70%	**71.33%**	**1.00**
21				Unbounded Steps	0.02%	23.68%	110.92%	150.70%	**71.33%**	**1.00**
22			Polynomial	Bounded Steps	0.01%	23.43%	109.70%	149.14%	**70.57%**	**1.00**
23				Unbounded Steps	0.05%	23.22%	108.67%	147.66%	**69.90%**	**1.00**

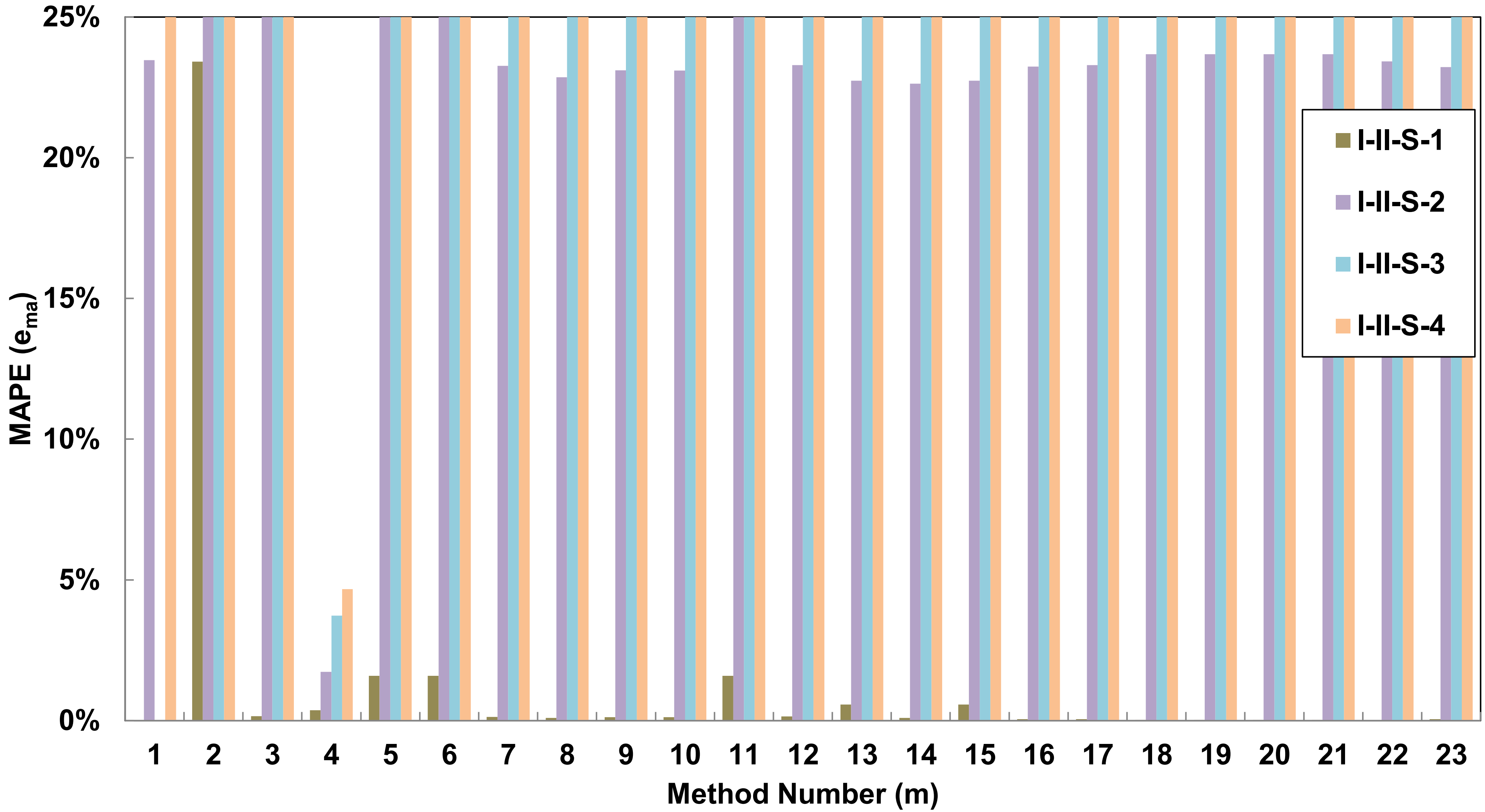


Figure G.1: MAPE of the Individual Test Set of Data Mining Models for Synchronous Flow Lines with the Training Set *I-II-S-1*

Table G.2: MAPE of the Individual Test Set of Data Mining Models for Synchronous Flow Lines with the Training Set *I-II-S-2*

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma}: Test Set $T_o, o=1$ *I-II-S-1*	Test Set $T_u, u=1$ *I-II-S-2*	Test Set $T_o, o=2$ *I-II-S-3*	Test Set $T_o, o=3$ *I-II-S-4*	μ_e	c_e
1	Decision Tree	Boosting			14.02%	17.93%	14.02%	112.51%	**39.62%**	**1.23**
2		Bootstrap Aggregating			39.18%	10.68%	287.51%	374.82%	**178.05%**	**1.02**
3	Neural Network	Feedforward			0.63%	0.12%	188.32%	239.85%	**107.23%**	**1.17**
4	Blumenfeld (1990)				0.37%	1.73%	3.73%	4.67%	**2.63%**	**0.74**
5	Multiple Linear Regression	Robust	Tukey's Bisquare		6.42%	26.26%	180.38%	243.08%	**114.04%**	**1.02**
6			Andrews		6.38%	26.31%	180.51%	243.09%	**114.07%**	**1.02**
7			Cauchy M-estimators by Moore		8.36%	24.70%	175.83%	238.43%	**111.83%**	**1.01**
8			Fair by Rey		18.77%	20.69%	147.06%	213.57%	**100.02%**	**0.97**
9			Huber		13.06%	22.35%	162.06%	225.65%	**105.78%**	**0.99**
10			Logistic Regression		13.99%	21.99%	158.76%	221.65%	**104.10%**	**0.99**
11			Hinch and Talwar		9.00%	24.91%	179.98%	249.28%	**115.79%**	**1.02**
12			Holland and Welsch		7.11%	25.71%	179.38%	242.91%	**113.78%**	**1.02**
13		Regularisation	Lasso		18.75%	19.97%	96.09%	137.84%	**68.16%**	**0.86**
14			Ridge Regression		22.19%	18.89%	113.93%	178.92%	**83.48%**	**0.93**
15			Elastic Nets		19.49%	19.42%	97.37%	123.34%	**64.91%**	**0.82**
16		Stepwise	Linear with Interactions of Linear Terms	Bounded Steps	1.42%	0.88%	59.12%	88.12%	**37.39%**	**1.16**
17				Unbounded Steps	1.41%	0.88%	61.31%	90.59%	**38.55%**	**1.16**
18			Linear and Squared Terms	Bounded Steps	22.58%	18.62%	116.30%	152.89%	**77.60%**	**0.87**
19				Unbounded Steps	1.56%	0.77%	61.62%	91.40%	**38.84%**	**1.16**
20			Linear and Squared Terms including Interactions	Bounded Steps	1.57%	0.77%	61.63%	91.44%	**38.85%**	**1.16**
21				Unbounded Steps	1.59%	0.76%	64.01%	97.93%	**41.07%**	**1.17**
22			Polynomial	Bounded Steps	0.17%	0.08%	62.35%	93.35%	**38.99%**	**1.20**
23				Unbounded Steps	0.15%	0.09%	64.35%	100.84%	**41.36%**	**1.21**

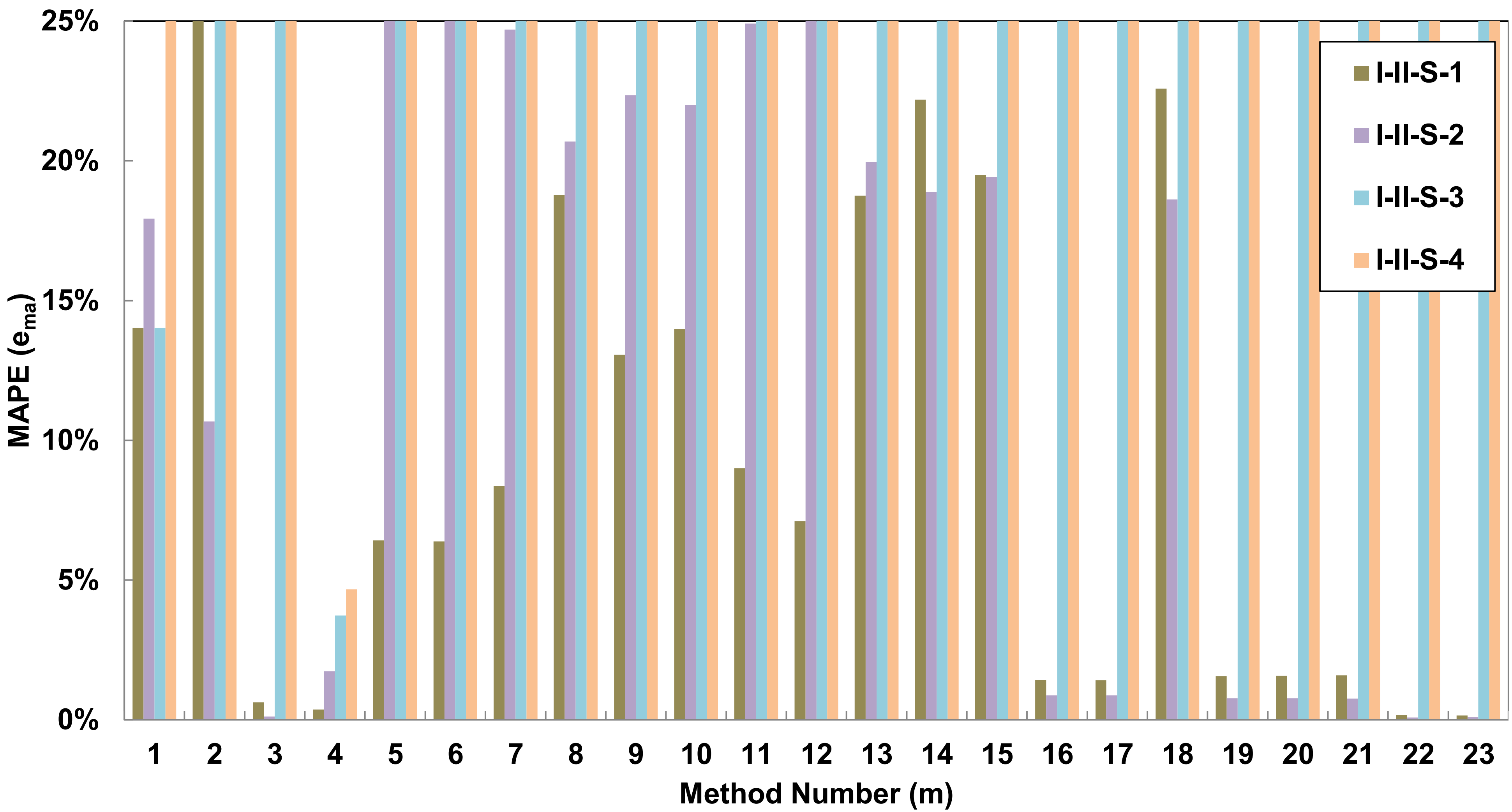


Figure G.2: MAPE of the Individual Test Set of Data Mining Models for Synchronous Flow Lines with the Training Set *I-II-S-2*

Table G.3: MAPE of the Individual Test Set of Data Mining Models for Synchronous Flow Lines with the Training Set *I-II-S-3*

					$\mathbf{e_{ma}}$					
	Method				**Test Set T$_u$,** u=1	**Test Set T$_u$,** u=2	**Test Set T$_u$,** u=3	**Test Set T$_o$,** o=1		
Number (m)	**Class I**	**Class II**	**Class III**	**Class IV**	***I-II-S-1***	***I-II-S-2***	***I-II-S-3***	***I-II-S-4***	μ_e	c_e
1	Decision Tree	Boosting			45.50%	25.64%	45.50%	26.01%	**35.66%**	**0.32**
2		Bootstrap Aggregating			45.36%	25.73%	7.70%	32.16%	**27.74%**	**0.56**
3	Neural Network	Feedforward			7.21%	5.66%	0.21%	1.87%	**3.74%**	**0.87**
4	Blumenfeld (1990)				0.37%	1.73%	3.73%	4.67%	**2.63%**	**0.74**
5	Multiple Linear Regression	Robust	Tukey's Bisquare		6.96%	12.27%	10.26%	14.26%	**10.94%**	**0.28**
6			Andrews		6.94%	12.27%	10.28%	14.22%	**10.93%**	**0.28**
7			Cauchy M-estimators by Moore		10.90%	12.73%	10.34%	17.76%	**12.93%**	**0.26**
8			Fair by Rey		14.23%	13.50%	11.09%	19.87%	**14.67%**	**0.25**
9			Huber		13.51%	13.36%	10.68%	19.03%	**14.15%**	**0.25**
10			Logistic Regression		13.58%	13.31%	10.75%	19.20%	**14.21%**	**0.25**
11			Hinch and Talwar		7.01%	12.25%	10.22%	13.92%	**10.85%**	**0.27**
12			Holland and Welsch		7.11%	12.24%	10.16%	14.55%	**11.02%**	**0.29**
13		Regularisation	Lasso		16.13%	14.69%	13.24%	23.67%	**16.93%**	**0.27**
14			Ridge Regression		16.14%	14.69%	13.24%	23.67%	**16.94%**	**0.27**
15			Elastic Nets		16.13%	14.69%	13.24%	23.67%	**16.93%**	**0.27**
16		Stepwise	Linear with Interactions of Linear Terms	Bounded Steps	2.72%	1.30%	0.83%	1.63%	**1.62%**	**0.50**
17				Unbounded Steps	2.72%	1.30%	0.83%	1.63%	**1.62%**	**0.50**
18			Linear and Squared Terms	Bounded Steps	16.18%	14.79%	13.16%	23.62%	**16.94%**	**0.27**
19				Unbounded Steps	2.65%	1.36%	0.79%	1.61%	**1.60%**	**0.49**
20			Linear and Squared Terms including Interactions	Bounded Steps	2.65%	1.37%	0.79%	1.61%	**1.61%**	**0.48**
21				Unbounded Steps	2.65%	1.36%	0.79%	1.61%	**1.60%**	**0.49**
22			Polynomial	Bounded Steps	0.23%	0.19%	0.21%	0.18%	**0.20%**	**0.11**
23				Unbounded Steps	3.51%	3.97%	0.21%	1.18%	**2.22%**	**0.82**

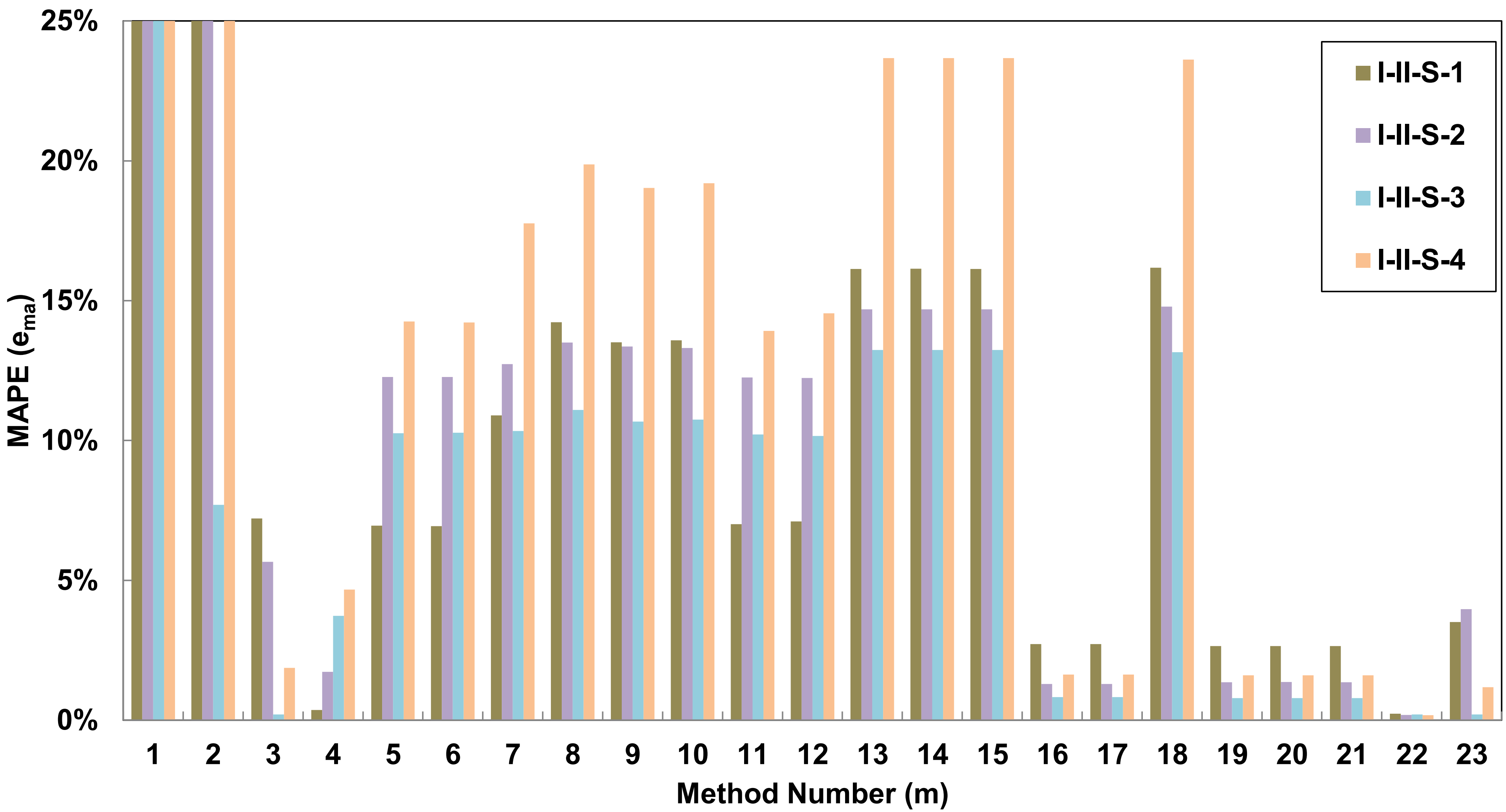


Figure G.3: MAPE of the Individual Test Set of Data Mining Models for Synchronous Flow Lines with the Training Set *I-II-S-3*

Table G.4: MAPE of the Individual Test Set of Data Mining Models for Synchronous Flow Lines with the Training Set *I-II-S-4*

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ *I-II-S-1*	Test Set $T_o, o=2$ *I-II-S-2*	Test Set $T_o, o=3$ *I-II-S-3*	Test Set $T_u, u=1$ *I-II-S-4*	μ_e	c_e
1	Decision Tree	Boosting			14.01%	15.94%	14.01%	26.88%	**17.71%**	**0.35**
2	Decision Tree	Bootstrap Aggregating			24.44%	15.28%	12.36%	30.33%	**20.60%**	**0.40**
3	Neural Network	Feedforward			159.04%	364.11%	730.29%	0.20%	**313.41%**	**1.01**
4	Blumenfeld (1990)				0.37%	1.73%	3.73%	4.67%	**2.63%**	**0.74**
5	Multiple Linear Regression	Robust	Tukey's Bisquare		12.10%	18.93%	20.08%	12.27%	**15.85%**	**0.27**
6	Multiple Linear Regression	Robust	Andrews		12.08%	18.82%	19.98%	12.24%	**15.78%**	**0.27**
7	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		13.22%	19.42%	18.44%	12.75%	**15.96%**	**0.22**
8	Multiple Linear Regression	Robust	Fair by Rey		15.58%	17.50%	13.35%	16.40%	**15.71%**	**0.11**
9	Multiple Linear Regression	Robust	Huber		14.50%	18.29%	14.97%	14.58%	**15.59%**	**0.12**
10	Multiple Linear Regression	Robust	Logistic Regression		14.59%	18.08%	14.53%	14.83%	**15.51%**	**0.11**
11	Multiple Linear Regression	Robust	Hinch and Talwar		17.18%	16.77%	13.17%	12.92%	**15.01%**	**0.15**
12	Multiple Linear Regression	Robust	Holland and Welsch		12.40%	19.19%	19.92%	12.31%	**15.96%**	**0.26**
13	Multiple Linear Regression	Regularisation	Lasso		13.70%	14.45%	15.58%	26.62%	**17.59%**	**0.35**
14	Multiple Linear Regression	Regularisation	Ridge Regression		13.87%	15.40%	15.50%	26.96%	**17.93%**	**0.34**
15	Multiple Linear Regression	Regularisation	Elastic Nets		13.70%	14.45%	15.58%	26.62%	**17.59%**	**0.35**
16	Multiple Linear Regression	Stepwise	Linear with Interactions of Linear Terms	Bounded Steps	3.64%	3.57%	2.57%	1.28%	**2.77%**	**0.40**
17	Multiple Linear Regression	Stepwise	Linear with Interactions of Linear Terms	Unbounded Steps	3.64%	3.57%	2.57%	1.28%	**2.77%**	**0.40**
18	Multiple Linear Regression	Stepwise	Linear and Squared Terms	Bounded Steps	13.91%	15.54%	15.47%	26.87%	**17.95%**	**0.33**
19	Multiple Linear Regression	Stepwise	Linear and Squared Terms	Unbounded Steps	3.60%	3.63%	2.56%	1.25%	**2.76%**	**0.41**
20	Multiple Linear Regression	Stepwise	Linear and Squared Terms including Interactions	Bounded Steps	3.60%	3.63%	2.56%	1.25%	**2.76%**	**0.41**
21	Multiple Linear Regression	Stepwise	Linear and Squared Terms including Interactions	Unbounded Steps	3.60%	3.63%	2.56%	1.25%	**2.76%**	**0.41**
22	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	0.60%	0.55%	0.41%	0.16%	**0.43%**	**0.46**
23	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	2.59%	2.55%	1.38%	0.18%	**1.68%**	**0.68**

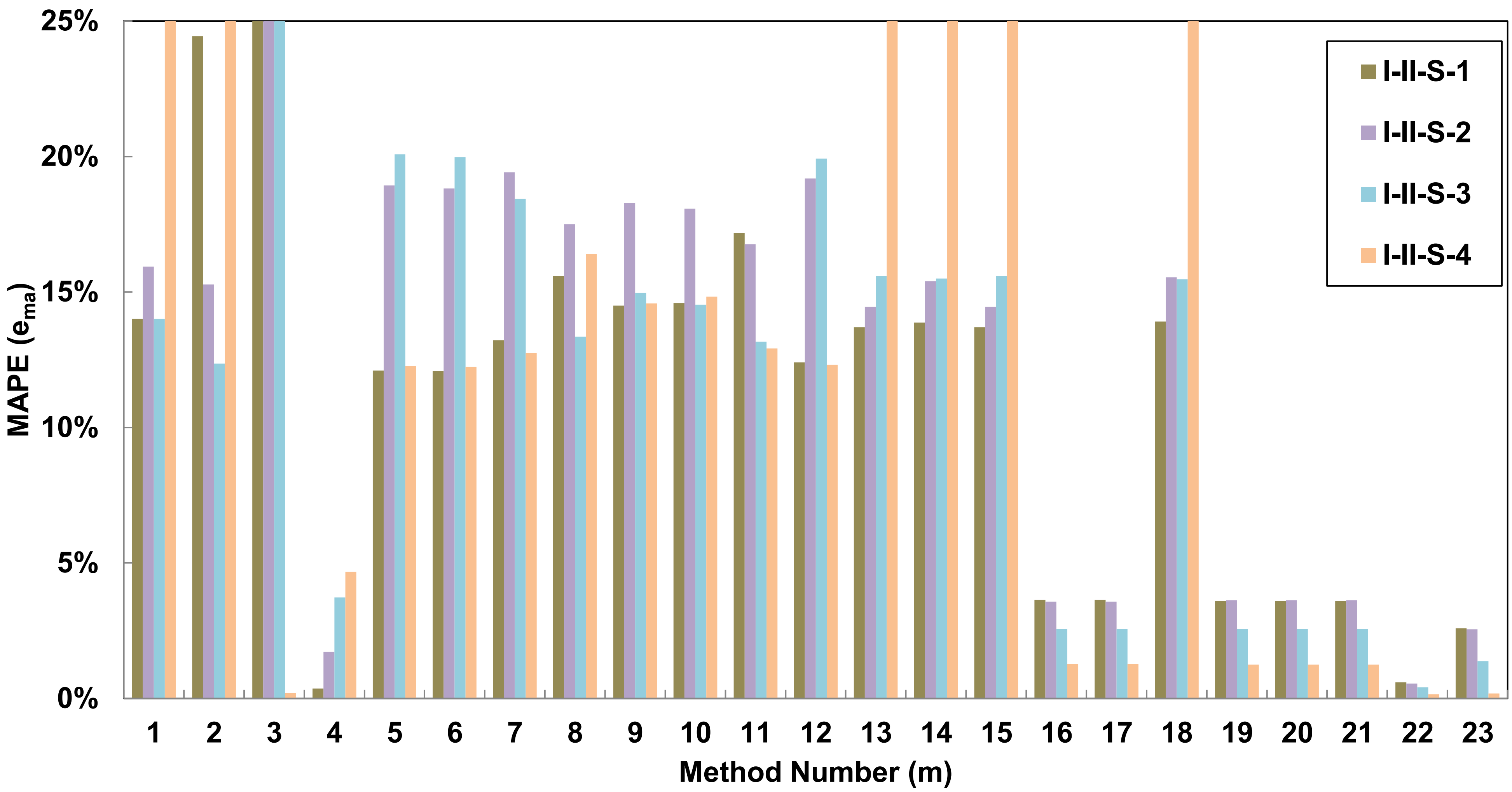


Figure G.4: MAPE of the Individual Test Set of Data Mining Models for Synchronous Flow Lines with the Training Set I-II-S-4

APPENDIX H – Phase III – MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines

Refer to Section 6.4.2.1.4 and 6.4.2.2.2

Table H.1: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-1*

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_u, u=1 *I-II-A-1*	e_{ma} Test Set T_o, o=1 *I-II-A-2*	e_{ma} Test Set T_o, o=2 *I-II-A-3*	e_{ma} Test Set T_o, o=3 *I-II-A-4*
1	Decision Tree	Boosting			4.61%	363.29%	3.61%	352.28%
2	Decision Tree	Bootstrap Aggregating			19.34%	23.96%	28.88%	543.59%
3	Neural Network	Feedforward			2.84%	241.25%	221.85%	1005.53%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		3.45%	25.53%	43.53%	78.93%
7	Multiple Linear Regression	Robust	Andrews		3.45%	25.54%	43.55%	78.95%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		3.49%	20.47%	33.12%	56.16%
9	Multiple Linear Regression	Robust	Fair by Rey		3.54%	12.00%	19.92%	43.82%
10	Multiple Linear Regression	Robust	Huber		3.41%	17.52%	29.35%	54.37%
11	Multiple Linear Regression	Robust	Logistic Regression		3.42%	16.51%	27.60%	51.56%
12	Multiple Linear Regression	Robust	Hinch and Talwar		3.45%	25.73%	43.88%	79.55%
13	Multiple Linear Regression	Robust	Holland and Welsch		3.45%	25.43%	43.34%	78.55%
14	Multiple Linear Regression	Regularisation	Lasso		4.78%	5.09%	4.33%	47.52%
15	Multiple Linear Regression	Regularisation	Ridge Regression		4.81%	5.03%	6.95%	53.62%
16	Multiple Linear Regression	Regularisation	Elastic Nets		4.81%	5.03%	6.95%	53.62%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	3.79%	22.09%	40.29%	124.39%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	3.77%	11.20%	18.29%	70.94%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	4.62%	24.13%	42.27%	121.87%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	3.37%	22.63%	39.96%	106.65%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	3.41%	22.90%	42.27%	126.87%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	3.45%	11.66%	19.24%	65.94%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	2.94%	7.72%	13.99%	35.35%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	2.82%	3.94%	5.22%	18.02%

Table H.1: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-1* (*cont.*)

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o,o=4 *I-II-A-5*	Test Set T_o,o=5 *I-II-A-6*	Test Set T_o,o=6 *I-II-A-7*	Test Set T_o,o=7 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			758.87%	521.22%	925.60%	5013.43%	**992.86%**	**1.67**
2	Decision Tree	Bootstrap Aggregating			1070.62%	736.93%	1305.15%	6901.48%	**1328.74%**	**1.73**
3	Neural Network	Feedforward			2854.79%	177.95%	371.06%	2246.42%	**890.21%**	**1.21**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		326.37%	3054.48%	5145.77%	25577.17%	**4281.90%**	**2.06**
7	Multiple Linear Regression	Robust	Andrews		326.48%	3055.38%	5147.29%	25584.75%	**4283.17%**	**2.06**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		230.92%	2177.65%	3666.66%	18208.46%	**3049.62%**	**2.06**
9	Multiple Linear Regression	Robust	Fair by Rey		139.08%	1492.66%	2503.18%	12420.87%	**2079.38%**	**2.06**
10	Multiple Linear Regression	Robust	Huber		200.92%	2078.31%	3494.49%	17354.26%	**2904.08%**	**2.06**
11	Multiple Linear Regression	Robust	Logistic Regression		186.96%	1962.39%	3298.35%	16377.69%	**2740.56%**	**2.06**
12	Multiple Linear Regression	Robust	Hinch and Talwar		329.56%	3079.40%	5187.97%	25787.37%	**4317.11%**	**2.06**
13	Multiple Linear Regression	Robust	Holland and Welsch		324.55%	3040.21%	5121.65%	25457.02%	**4261.78%**	**2.06**
14	Multiple Linear Regression	Regularisation	Lasso		96.88%	58.17%	119.28%	646.79%	**122.86%**	**1.76**
15	Multiple Linear Regression	Regularisation	Ridge Regression		110.10%	890.90%	1482.40%	7360.92%	**1239.34%**	**2.04**
16	Multiple Linear Regression	Regularisation	Elastic Nets		110.10%	890.90%	1482.40%	7360.92%	**1239.34%**	**2.04**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	455.28%	2696.23%	5067.29%	25453.85%	**4232.90%**	**2.07**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	213.85%	403.84%	737.45%	3727.06%	**648.30%**	**1.96**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	424.75%	2685.53%	4580.83%	22850.81%	**3841.85%**	**2.05**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	387.54%	2278.94%	4636.30%	22982.92%	**3807.29%**	**2.08**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	484.28%	3059.77%	5587.57%	28401.78%	**4716.11%**	**2.07**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	209.32%	442.09%	774.00%	3914.45%	**680.02%**	**1.96**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	125.84%	624.97%	1210.87%	5973.31%	**999.37%**	**2.06**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	58.26%	107.88%	199.55%	1024.61%	**177.54%**	**1.97**

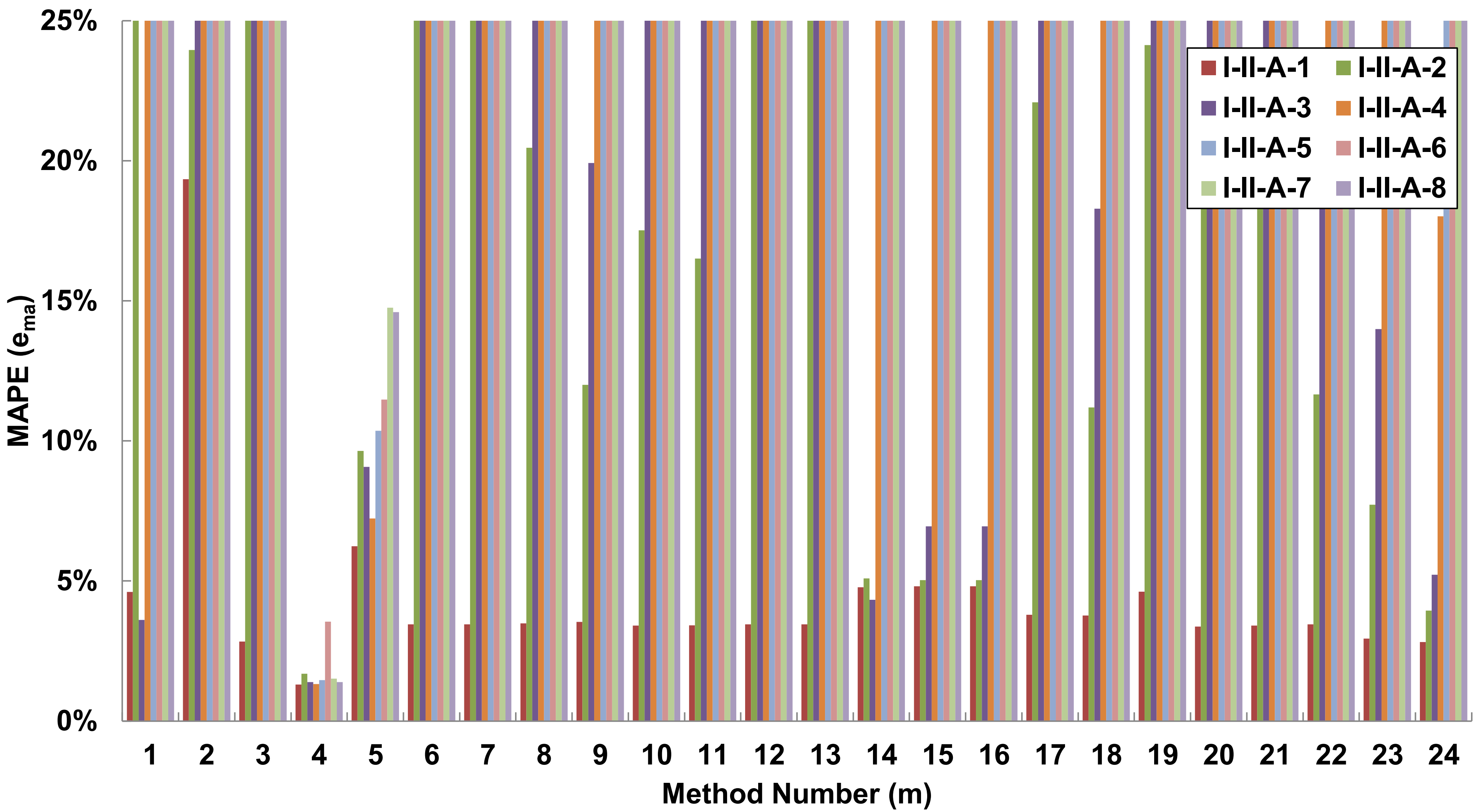


Figure H.1: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-1*

Table H.2: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-2*

	Method				e_{ma}			
					Test Set $T_o, o=1$	Test Set $T_u, u=1$	Test Set $T_o, o=2$	Test Set $T_o, o=3$
Number (m)	Class I	Class II	Class III	Class IV	*I-II-A-1*	*I-II-A-2*	*I-II-A-3*	*I-II-A-4*
1	Decision Tree	Boosting			4.95%	349.93%	3.17%	339.61%
2		Bootstrap Aggregating			9.32%	7.01%	6.45%	390.22%
3	Neural Network	Feedforward						
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		3.61%	4.04%	2.60%	32.10%
7			Andrews		3.61%	4.05%	2.60%	32.10%
8			Cauchy M-estimators by Moore		3.80%	3.77%	2.37%	35.93%
9			Fair by Rey		4.10%	3.80%	2.58%	39.05%
10			Huber		3.99%	3.77%	2.51%	37.92%
11			Logistic Regression		4.00%	3.77%	2.52%	37.98%
12			Hinch and Talwar		3.64%	4.05%	2.59%	32.76%
13			Holland and Welsch		3.60%	4.01%	2.55%	32.24%
14		Regularisation	Lasso					
15			Ridge Regression					
16			Elastic Nets					
17		Stepwise	Interaction	Bounded Steps	3.54%	3.32%	2.33%	20.60%
18				Unbounded Steps	3.47%	3.32%	2.42%	16.35%
19			Purequadratic	Bounded Steps	5.09%	3.96%	2.97%	48.05%
20				Unbounded Steps	3.63%	3.33%	2.44%	15.31%
21			Quadratic	Bounded Steps	3.60%	3.33%	2.25%	20.76%
22				Unbounded Steps	3.57%	3.33%	2.38%	15.86%
23			Polynomial	Bounded Steps	3.09%	3.28%	2.22%	10.01%
24				Unbounded Steps				

Table H.2: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-2* (*cont.*)

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o,o=4 *I-II-A-5*	Test Set T_o,o=5 *I-II-A-6*	Test Set T_o,o=6 *I-II-A-7*	Test Set T_o,o=7 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			746.56%	511.98%	910.84%	4939.76%	**975.85%**	**1.67**
2		Bootstrap Aggregating			827.20%	566.21%	1010.10%	5432.97%	**1031.19%**	**1.77**
3	Neural Network	Feedforward								
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		63.60%	42.69%	94.72%	515.13%	**94.81%**	**1.82**
7			Andrews		63.58%	42.82%	94.86%	515.59%	**94.90%**	**1.82**
8			Cauchy M-estimators by Moore		74.18%	37.82%	94.10%	519.53%	**96.44%**	**1.81**
9			Fair by Rey		80.24%	60.84%	132.30%	700.09%	**127.88%**	**1.84**
10			Huber		78.30%	48.40%	112.15%	604.21%	**111.41%**	**1.82**
11			Logistic Regression		78.41%	48.34%	111.94%	602.65%	**111.20%**	**1.82**
12			Hinch and Talwar		64.93%	42.93%	96.24%	523.12%	**96.28%**	**1.82**
13			Holland and Welsch		64.12%	41.32%	93.63%	511.29%	**94.10%**	**1.82**
14		Regularisation	Lasso							
15			Ridge Regression							
16			Elastic Nets							
17		Stepwise	Interaction	Bounded Steps	15.49%	8.18%	18.10%	101.27%	**21.60%**	**1.53**
18				Unbounded Steps	13.08%	48.15%	99.20%	510.80%	**87.10%**	**2.00**
19			Purequadratic	Bounded Steps	91.38%	49.33%	109.78%	598.32%	**113.61%**	**1.76**
20				Unbounded Steps	9.71%	59.09%	101.88%	515.28%	**88.83%**	**1.98**
21			Quadratic	Bounded Steps	11.55%	13.83%	24.05%	123.30%	**25.33%**	**1.60**
22				Unbounded Steps	9.73%	56.63%	100.23%	513.61%	**88.17%**	**1.99**
23			Polynomial	Bounded Steps	8.32%	13.41%	23.99%	115.26%	**22.45%**	**1.70**
24				Unbounded Steps						

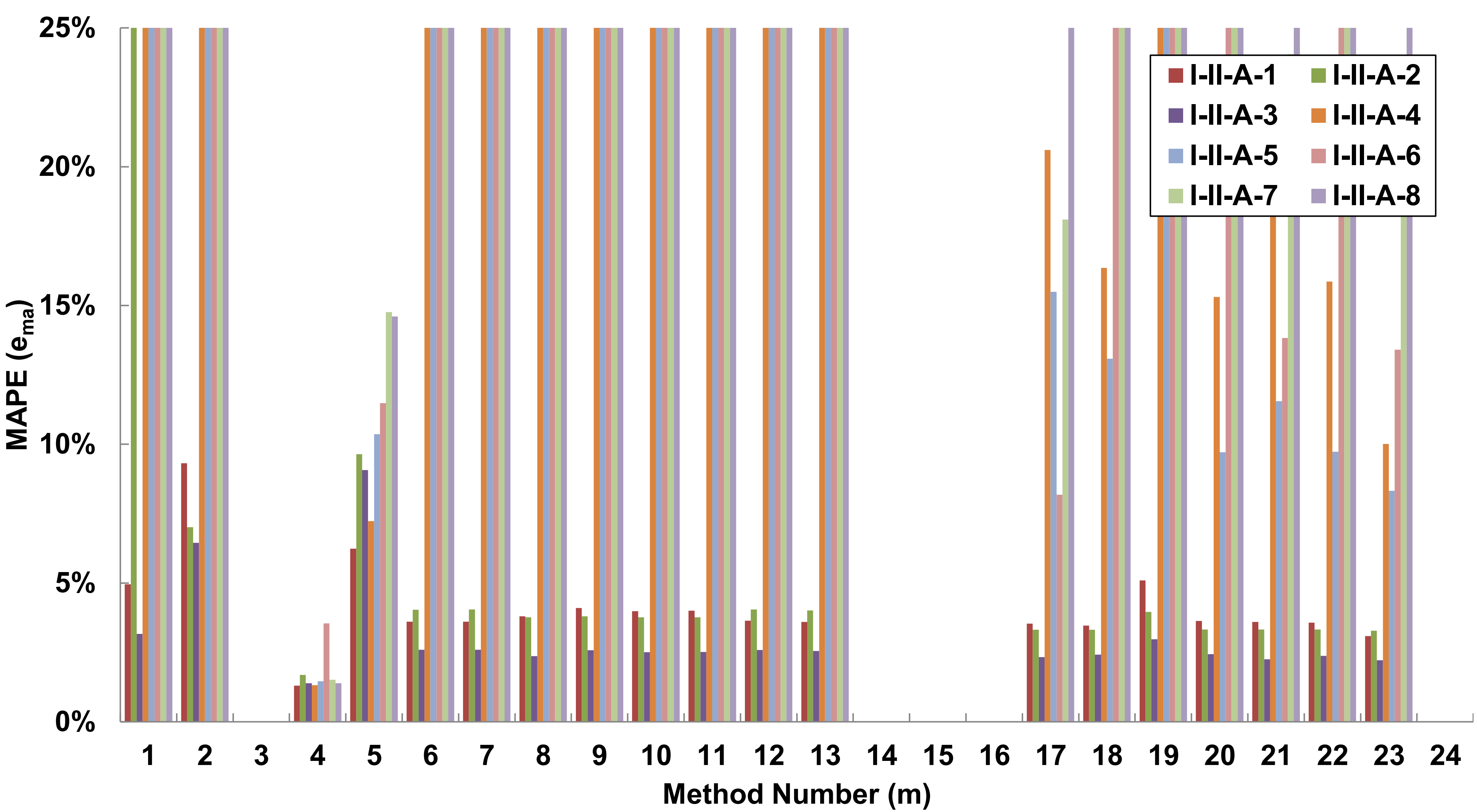


Figure H.2: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-2*

Table H.3: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-3*

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ *I-II-A-1*	Test Set $T_o, o=2$ *I-II-A-2*	Test Set $T_u, u=1$ *I-II-A-3*	Test Set $T_o, o=3$ *I-II-A-4*
1	Decision Tree	Boosting			11.03%	388.59%	1.79%	337.20%
2		Bootstrap Aggregating			22.51%	11.74%	5.89%	368.71%
3	Neural Network	Feedforward			16.78%	8.05%	2.36%	174.26%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		25.82%	10.93%	2.43%	127.40%
7			Andrews		25.84%	10.94%	2.43%	127.45%
8			Cauchy M-estimators by Moore		21.39%	9.45%	2.20%	108.23%
9			Fair by Rey		18.17%	8.15%	2.24%	74.94%
10			Huber		18.06%	8.13%	2.22%	81.67%
11			Logistic Regression		18.11%	8.15%	2.22%	80.54%
12			Hinch and Talwar		24.42%	10.55%	2.35%	125.86%
13			Holland and Welsch		25.68%	10.89%	2.42%	127.11%
14		Regularisation	Lasso		5.02%	4.15%	2.42%	40.27%
15			Ridge Regression		5.06%	4.19%	2.37%	40.44%
16			Elastic Nets		5.02%	4.15%	2.42%	40.27%
17		Stepwise	Interaction	Bounded Steps	4.17%	3.55%	2.02%	17.93%
18				Unbounded Steps	4.56%	3.69%	2.06%	23.11%
19			Purequadratic	Bounded Steps	5.02%	4.15%	2.42%	40.27%
20				Unbounded Steps	4.46%	3.67%	1.98%	18.80%
21			Quadratic	Bounded Steps	4.17%	3.55%	2.02%	17.93%
22				Unbounded Steps	4.46%	3.67%	1.98%	18.80%
23			Polynomial	Bounded Steps	4.17%	3.55%	2.02%	17.93%
24				Unbounded Steps	4.56%	3.69%	2.06%	23.11%

Table H.3: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-3* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=4 *I-II-A-5*	Test Set T_o, o=5 *I-II-A-6*	Test Set T_o, o=6 *I-II-A-7*	Test Set T_o, o=7 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			746.24%	512.55%	911.41%	4942.30%	**981.39%**	**1.66**
2		Bootstrap Aggregating			806.67%	547.31%	988.36%	5323.05%	**1009.28%**	**1.77**
3	Neural Network	Feedforward			427.73%	286.69%	536.54%	3010.94%	**557.92%**	**1.81**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		69.47%	1827.22%	3062.08%	15196.90%	**2540.28%**	**2.06**
7			Andrews		69.40%	1828.97%	3065.05%	15211.72%	**2542.73%**	**2.06**
8			Cauchy M-estimators by Moore		69.90%	1443.33%	2412.80%	11966.65%	**2004.24%**	**2.06**
9			Fair by Rey		73.67%	1143.35%	1911.91%	9495.91%	**1591.04%**	**2.06**
10			Huber		70.34%	1159.09%	1936.47%	9609.53%	**1610.69%**	**2.06**
11			Logistic Regression		70.60%	1157.74%	1934.58%	9601.65%	**1609.20%**	**2.06**
12			Hinch and Talwar		78.87%	1702.66%	2850.08%	14138.42%	**2366.65%**	**2.06**
13			Holland and Welsch		70.07%	1815.46%	3042.10%	15097.26%	**2523.87%**	**2.06**
14		Regularisation	Lasso		83.40%	49.37%	102.41%	567.30%	**106.79%**	**1.78**
15			Ridge Regression		83.40%	49.59%	102.49%	567.03%	**106.82%**	**1.78**
16			Elastic Nets		83.40%	49.37%	102.41%	567.30%	**106.79%**	**1.78**
17		Stepwise	Interaction	Bounded Steps	40.08%	26.20%	54.34%	269.15%	**52.18%**	**1.72**
18				Unbounded Steps	51.85%	35.97%	67.45%	341.33%	**66.25%**	**1.72**
19			Purequadratic	Bounded Steps	83.40%	49.37%	102.41%	567.30%	**106.79%**	**1.78**
20				Unbounded Steps	41.37%	30.41%	53.60%	271.04%	**53.17%**	**1.69**
21			Quadratic	Bounded Steps	40.08%	26.20%	54.34%	269.15%	**52.18%**	**1.72**
22				Unbounded Steps	41.37%	30.41%	53.60%	271.04%	**53.17%**	**1.69**
23			Polynomial	Bounded Steps	40.08%	26.20%	54.34%	269.15%	**52.18%**	**1.72**
24				Unbounded Steps	51.85%	35.97%	67.45%	341.33%	**66.25%**	**1.72**

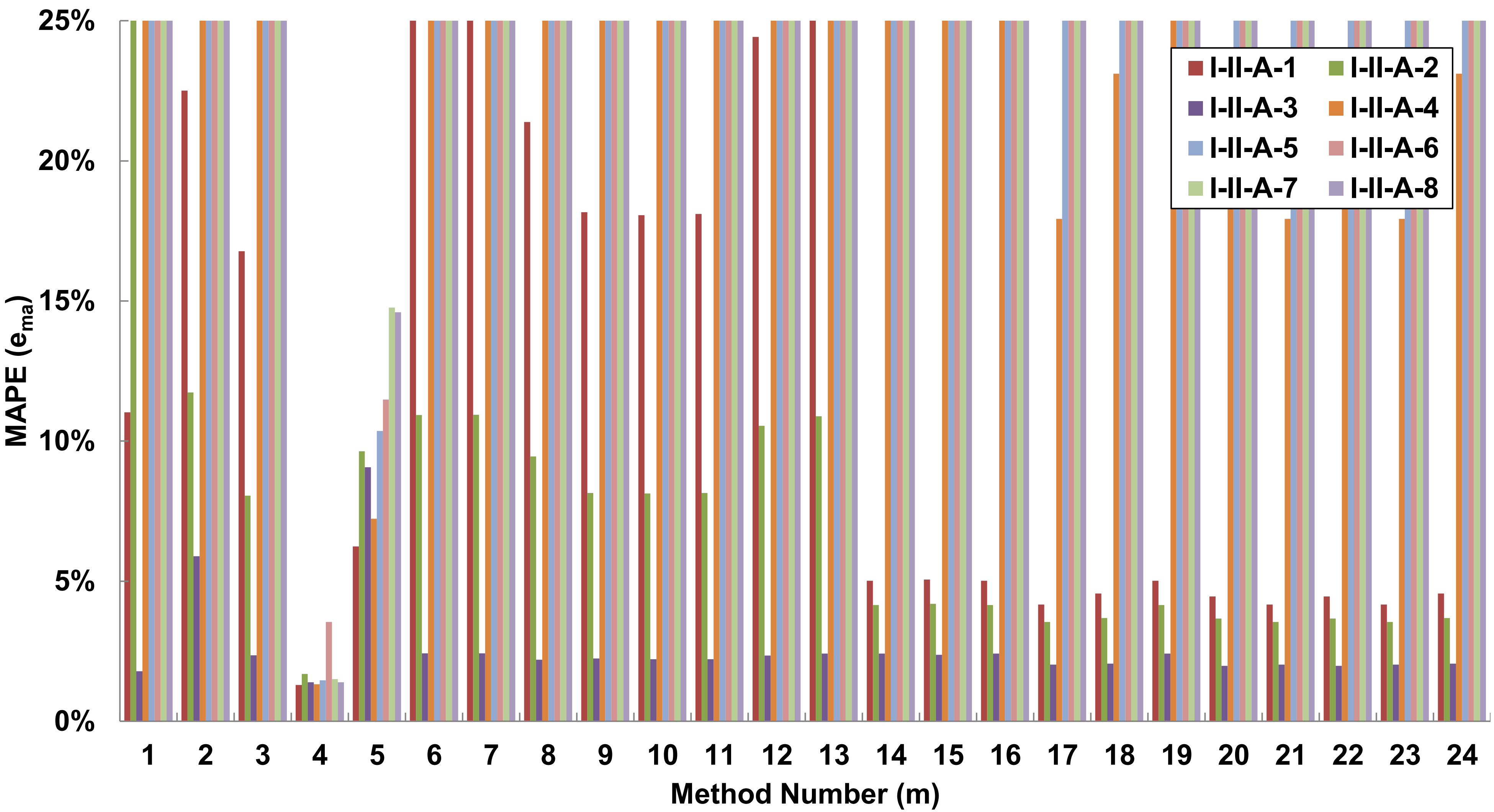


Figure H.3: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-3*

Table H.4: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-4*

Number (m)	Class I	Class II	Class III	Class IV	$\mathbf{e_{ma}}$ Test Set $T_o, o=1$ *I-II-A-1*	Test Set $T_o, o=2$ *I-II-A-2*	Test Set $T_o, o=3$ *I-II-A-3*	Test Set $T_u, u=1$ *I-II-A-4*
1	Decision Tree	Boosting			5.98%	30.03%	6.85%	3.54%
2		Bootstrap Aggregating			40.31%	41.63%	39.35%	19.77%
3	Neural Network	Feedforward			7.32%	5.25%	2.72%	3.31%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.28%	8.37%	7.70%	3.21%
7			Andrews		5.28%	8.37%	7.70%	3.21%
8			Cauchy M-estimators by Moore		5.28%	8.37%	7.69%	3.19%
9			Fair by Rey		5.18%	8.24%	7.27%	3.34%
10			Huber		5.01%	8.20%	7.21%	3.20%
11			Logistic Regression		5.01%	8.19%	7.21%	3.21%
12			Hinch and Talwar		5.28%	8.37%	7.69%	3.20%
13			Holland and Welsch		5.29%	8.38%	7.70%	3.21%
14		Regularisation	Lasso		5.31%	8.15%	7.12%	3.76%
15			Ridge Regression		5.34%	8.26%	7.28%	3.63%
16			Elastic Nets		5.31%	8.15%	7.12%	3.76%
17		Stepwise	Interaction	Bounded Steps	4.34%	6.99%	5.79%	2.87%
18				Unbounded Steps	4.12%	7.27%	6.18%	2.81%
19			Purequadratic	Bounded Steps	5.33%	8.24%	7.25%	3.67%
20				Unbounded Steps	4.19%	6.93%	5.85%	2.75%
21			Quadratic	Bounded Steps	4.27%	6.80%	5.59%	2.78%
22				Unbounded Steps	4.09%	7.16%	6.06%	2.79%
23			Polynomial	Bounded Steps	3.30%	4.08%	4.23%	2.52%
24				Unbounded Steps	376.42%	260.09%	135.97%	2.12%

Table H.4: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-4* (*cont.*)

	Method				e_{ma}					
					Test Set T_o, o=4	Test Set T_o, o=5	Test Set T_o, o=6	Test Set T_o, o=7		
Number (m)	Class I	Class II	Class III	Class IV	*I-II-A-5*	*I-II-A-6*	*I-II-A-7*	*I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			40.64%	3.63%	65.74%	726.35%	**110.35%**	**2.26**
2		Bootstrap Aggregating			55.14%	10.72%	83.96%	817.01%	**138.49%**	**1.99**
3	Neural Network	Feedforward			5.96%	6115.20%	10262.86%	51451.52%	**8481.77%**	**2.10**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.56%	4.31%	7.05%	80.79%	**15.28%**	**1.74**
7			Andrews		5.56%	4.31%	7.12%	81.44%	**15.37%**	**1.74**
8			Cauchy M-estimators by Moore		5.52%	4.28%	6.81%	79.30%	**15.06%**	**1.73**
9			Fair by Rey		4.98%	38.44%	72.90%	413.22%	**69.20%**	**2.04**
10			Huber		5.36%	24.62%	50.57%	307.37%	**51.44%**	**2.03**
11			Logistic Regression		5.28%	25.39%	51.77%	312.91%	**52.37%**	**2.03**
12			Hinch and Talwar		5.57%	4.26%	6.83%	78.93%	**15.02%**	**1.72**
13			Holland and Welsch		5.56%	4.33%	6.80%	77.92%	**14.90%**	**1.71**
14		Regularisation	Lasso		6.93%	4.88%	7.52%	82.12%	**15.72%**	**1.71**
15			Ridge Regression		5.88%	3.32%	7.67%	82.39%	**15.47%**	**1.75**
16			Elastic Nets		6.93%	4.88%	7.52%	82.12%	**15.72%**	**1.71**
17		Stepwise	Interaction	Bounded Steps	6.59%	2.93%	9.58%	90.21%	**16.16%**	**1.86**
18				Unbounded Steps	6.70%	171.58%	269.07%	1357.57%	**228.16%**	**2.05**
19			Purequadratic	Bounded Steps	6.02%	3.06%	7.53%	83.63%	**15.59%**	**1.77**
20				Unbounded Steps	5.59%	8.73%	9.16%	28.90%	**9.01%**	**0.92**
21			Quadratic	Bounded Steps	6.42%	2.47%	9.30%	88.21%	**15.73%**	**1.87**
22				Unbounded Steps	6.58%	176.65%	277.49%	1392.50%	**234.17%**	**2.05**
23			Polynomial	Bounded Steps	4.77%	3.49%	8.09%	52.12%	**10.33%**	**1.64**
24				Unbounded Steps	16.14%	143.66%	949.16%	14509.31%	**2049.11%**	**2.46**

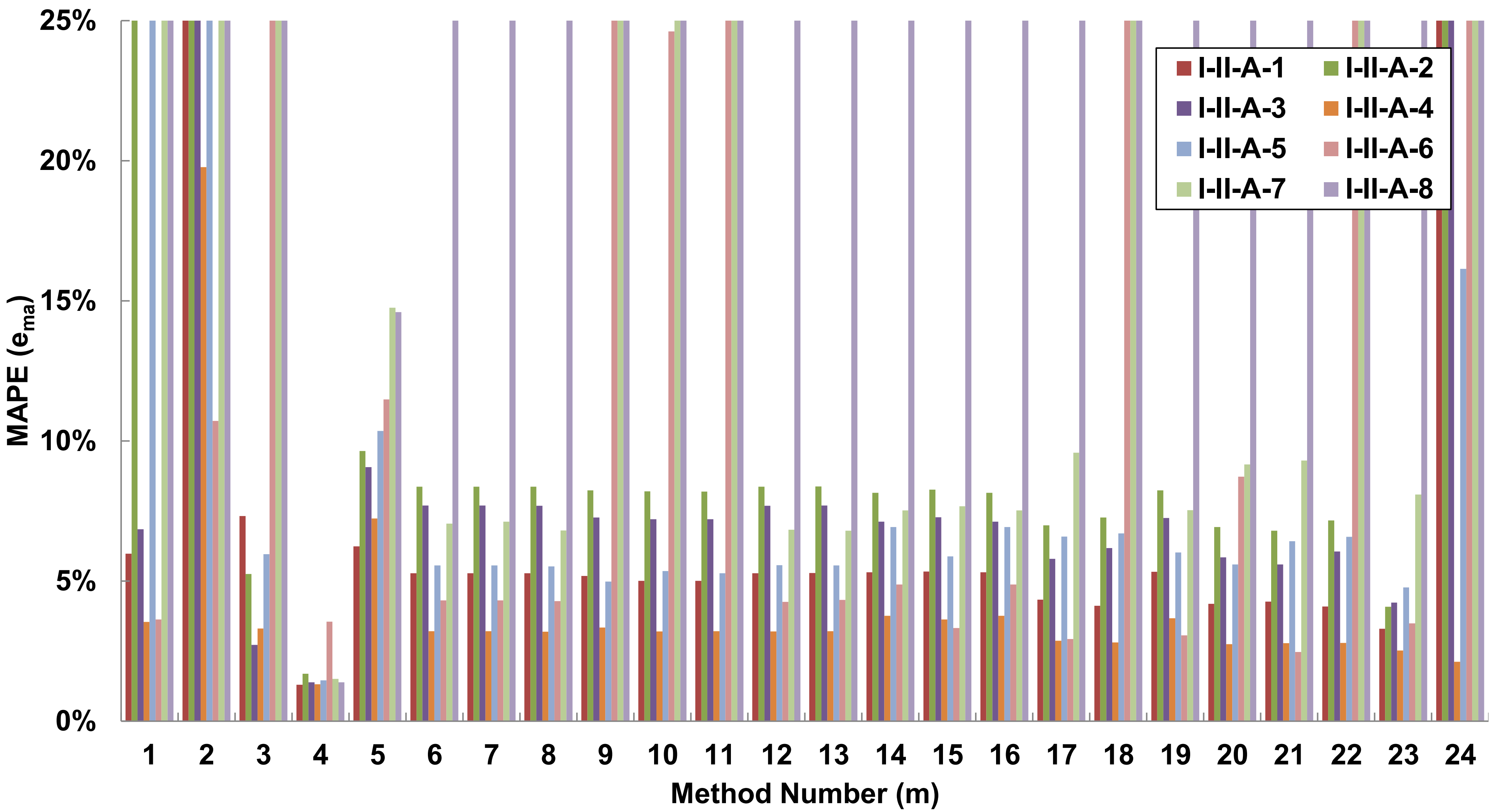


Figure H.4: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-4*

Table H.5: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-5*

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1 *I-II-A-1*	Test Set T_o, o=2 *I-II-A-2*	Test Set T_o, o=3 *I-II-A-3*	Test Set T_o, o=4 *I-II-A-4*
1	Decision Tree	Boosting			66.99%	49.54%	55.49%	9.74%
2		Bootstrap Aggregating			85.61%	82.23%	80.59%	25.80%
3	Neural Network	Feedforward			46.83%	35.11%	30.29%	8.38%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.88%	8.68%	8.19%	10.55%
7			Andrews		5.88%	8.68%	8.19%	10.57%
8			Cauchy M-estimators by Moore		5.49%	8.57%	8.05%	7.27%
9			Fair by Rey		5.50%	8.53%	7.55%	6.01%
10			Huber		5.45%	8.53%	7.65%	6.12%
11			Logistic Regression		5.45%	8.53%	7.65%	6.15%
12			Hinch and Talwar		5.88%	8.68%	8.18%	10.55%
13			Holland and Welsch		5.87%	8.68%	8.19%	10.50%
14		Regularisation	Lasso		6.07%	8.83%	7.84%	4.39%
15			Ridge Regression		6.07%	8.83%	7.84%	4.39%
16			Elastic Nets		6.07%	8.83%	7.84%	4.39%
17		Stepwise	Interaction	Bounded Steps	3.49%	3.91%	2.69%	3.40%
18				Unbounded Steps	3.49%	3.91%	2.69%	3.40%
19			Purequadratic	Bounded Steps	40.41%	29.52%	24.88%	6.63%
20				Unbounded Steps	3.53%	3.88%	2.65%	3.39%
21			Quadratic	Bounded Steps	3.50%	3.90%	2.68%	3.40%
22				Unbounded Steps	3.53%	3.88%	2.65%	3.39%
23			Polynomial	Bounded Steps	3.51%	3.89%	2.69%	3.39%
24				Unbounded Steps	36.70%	25.49%	22.03%	6.09%

Table H.5: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-5* (*cont.*)

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_u, u=1 *I-II-A-5*	Test Set T_o, o=5 *I-II-A-6*	Test Set T_o, o=6 *I-II-A-7*	Test Set T_o, o=7 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			3.22%	6.04%	3.70%	396.63%	**73.92%**	**1.80**
2	Decision Tree	Bootstrap Aggregating			9.48%	13.05%	17.53%	483.92%	**99.78%**	**1.59**
3	Neural Network	Feedforward			2.88%	3.49%	4.33%	214.35%	**43.21%**	**1.65**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		3.83%	197.30%	326.58%	1568.44%	**266.18%**	**2.03**
7	Multiple Linear Regression	Robust	Andrews		3.83%	197.73%	327.29%	1572.00%	**266.77%**	**2.03**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		3.39%	111.93%	182.03%	848.37%	**146.89%**	**1.98**
9	Multiple Linear Regression	Robust	Fair by Rey		3.38%	72.04%	115.08%	518.10%	**92.02%**	**1.92**
10	Multiple Linear Regression	Robust	Huber		3.38%	77.82%	124.64%	564.31%	**99.74%**	**1.93**
11	Multiple Linear Regression	Robust	Logistic Regression		3.38%	78.31%	125.47%	568.53%	**100.43%**	**1.94**
12	Multiple Linear Regression	Robust	Hinch and Talwar		3.83%	197.30%	326.57%	1568.41%	**266.18%**	**2.03**
13	Multiple Linear Regression	Robust	Holland and Welsch		3.82%	196.04%	324.44%	1557.76%	**264.41%**	**2.03**
14	Multiple Linear Regression	Regularisation	Lasso		3.57%	4.95%	3.26%	51.12%	**11.25%**	**1.44**
15	Multiple Linear Regression	Regularisation	Ridge Regression		3.57%	4.95%	3.26%	51.12%	**11.25%**	**1.44**
16	Multiple Linear Regression	Regularisation	Elastic Nets		3.57%	4.95%	3.26%	51.12%	**11.25%**	**1.44**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	2.85%	3.94%	3.73%	14.74%	**4.84%**	**0.83**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	2.85%	3.94%	3.73%	14.74%	**4.84%**	**0.83**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	3.31%	5.27%	3.33%	56.38%	**21.22%**	**0.94**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	2.80%	3.89%	3.69%	14.17%	**4.75%**	**0.81**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	2.83%	3.93%	3.70%	14.69%	**4.83%**	**0.83**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	2.80%	3.89%	3.69%	14.17%	**4.75%**	**0.81**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	2.81%	3.90%	3.70%	14.25%	**4.77%**	**0.81**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	2.80%	3.63%	3.78%	10.70%	**13.90%**	**0.91**

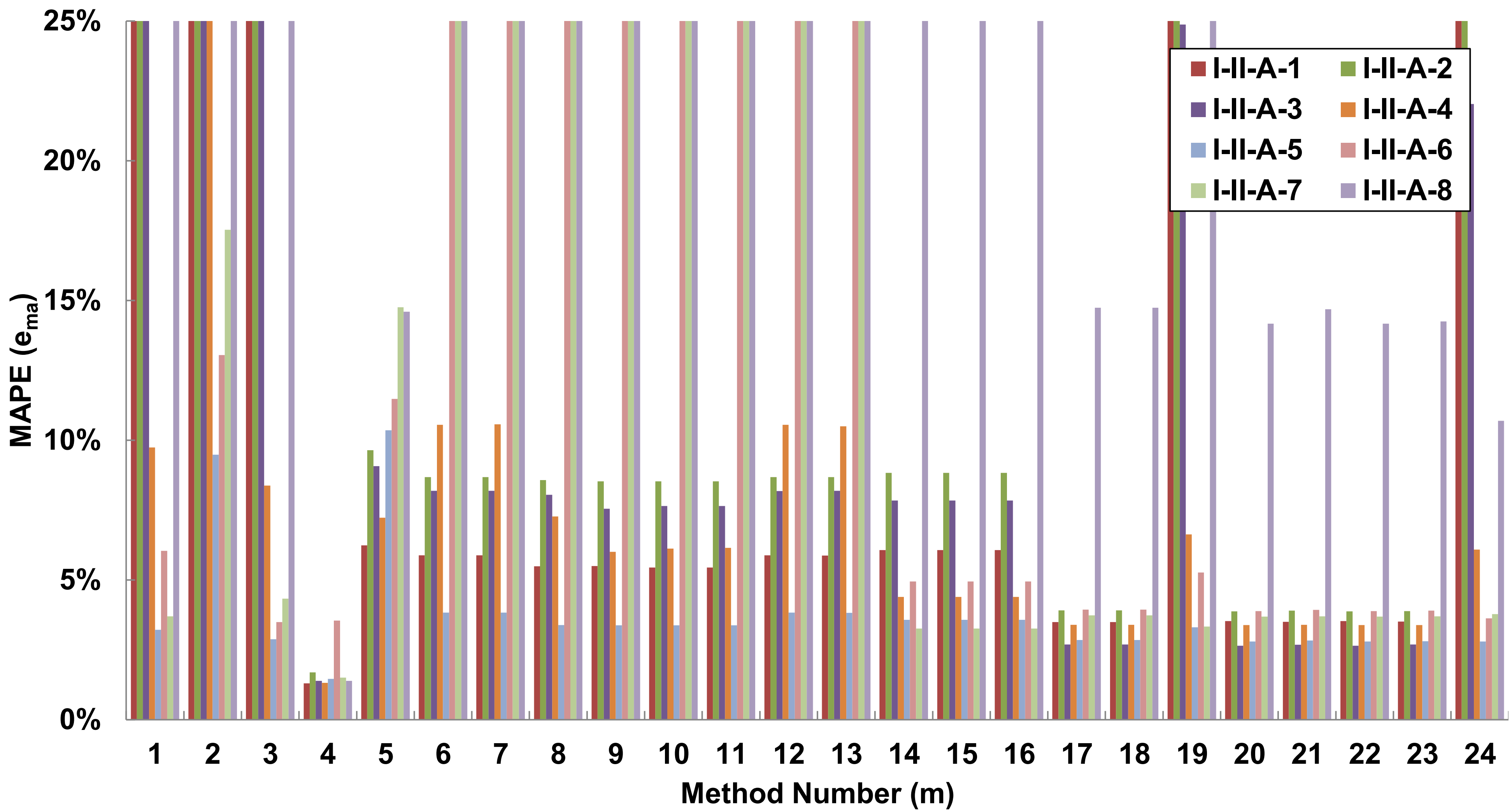


Figure H.5: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-5*

Table H.6: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-6*

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma}: Test Set T_o, o=1 *I-II-A-1*	Test Set T_o, o=2 *I-II-A-2*	Test Set T_o, o=3 *I-II-A-3*	Test Set T_o, o=4 *I-II-A-4*
1	Decision Tree	Boosting			87.80%	48.30%	83.59%	24.13%
2		Bootstrap Aggregating			88.37%	85.64%	84.34%	26.72%
3	Neural Network	Feedforward			94.07%	92.64%	92.00%	34.74%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		8.03%	9.67%	8.49%	19.68%
7			Andrews		8.10%	9.72%	8.54%	19.89%
8			Cauchy M-estimators by Moore		7.57%	9.37%	8.19%	17.79%
9			Fair by Rey		7.20%	9.13%	8.00%	16.72%
10			Huber		8.10%	9.71%	8.51%	18.35%
11			Logistic Regression		7.46%	9.30%	8.13%	17.37%
12			Hinch and Talwar		7.53%	9.32%	8.13%	16.90%
13			Holland and Welsch		7.80%	9.52%	8.34%	18.95%
14		Regularisation	Lasso		6.00%	8.33%	7.20%	4.73%
15			Ridge Regression		6.00%	8.33%	7.20%	4.73%
16			Elastic Nets		6.00%	8.33%	7.20%	4.73%
17		Stepwise	Interaction	Bounded Steps	5.28%	7.96%	6.91%	4.24%
18				Unbounded Steps	31.83%	32.53%	33.12%	12.31%
19			Purequadratic	Bounded Steps	4.87%	7.78%	6.73%	3.97%
20				Unbounded Steps	4227.10%	2741.21%	2294.17%	335.43%
21			Quadratic	Bounded Steps	4.89%	7.79%	6.74%	3.98%
22				Unbounded Steps	34.56%	35.37%	35.97%	13.01%
23			Polynomial	Bounded Steps	4.94%	7.80%	6.74%	4.01%
24				Unbounded Steps	5.28%	7.96%	6.91%	4.24%

Table H.6: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-6* (*cont.*)

Number (m)	Method: Class I	Method: Class II	Method: Class III	Method: Class IV	e_{ma}: Test Set T_o, o=5 *I-II-A-5*	Test Set T_u, u=1 *I-II-A-6*	Test Set T_o, o=6 *I-II-A-7*	Test Set T_o, o=7 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			39.28%	1.26%	62.42%	709.84%	**132.08%**	**1.78**
2		Bootstrap Aggregating			42.14%	4.10%	66.90%	731.82%	**141.25%**	**1.70**
3	Neural Network	Feedforward			33.47%	1.32%	49.38%	702.67%	**137.54%**	**1.68**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		34.16%	1.66%	9.67%	102.84%	**24.28%**	**1.37**
7			Andrews		34.45%	1.66%	9.67%	102.83%	**24.36%**	**1.37**
8			Cauchy M-estimators by Moore		31.21%	1.65%	9.51%	102.02%	**23.41%**	**1.41**
9			Fair by Rey		29.79%	1.65%	9.41%	101.34%	**22.91%**	**1.43**
10			Huber		31.37%	1.65%	9.58%	102.58%	**23.73%**	**1.39**
11			Logistic Regression		30.58%	1.65%	9.46%	101.74%	**23.21%**	**1.42**
12			Hinch and Talwar		29.50%	1.70%	9.87%	104.05%	**23.38%**	**1.44**
13			Holland and Welsch		33.11%	1.65%	9.61%	102.57%	**23.94%**	**1.39**
14		Regularisation	Lasso		8.77%	1.70%	9.39%	101.23%	**18.42%**	**1.82**
15			Ridge Regression		8.77%	1.70%	9.39%	101.23%	**18.42%**	**1.82**
16			Elastic Nets		8.77%	1.70%	9.39%	101.23%	**18.42%**	**1.82**
17		Stepwise	Interaction	Bounded Steps	8.48%	1.58%	9.09%	98.42%	**17.75%**	**1.84**
18				Unbounded Steps	15.68%	1.49%	26.09%	288.60%	**55.21%**	**1.72**
19			Purequadratic	Bounded Steps	8.56%	1.32%	9.27%	100.14%	**17.83%**	**1.87**
20				Unbounded Steps	76.88%	1.32%	134.84%	3303.94%	**1639.36%**	**1.04**
21			Quadratic	Bounded Steps	8.54%	1.31%	9.26%	99.97%	**17.81%**	**1.87**
22				Unbounded Steps	17.02%	1.22%	28.32%	314.09%	**59.95%**	**1.73**
23			Polynomial	Bounded Steps	8.52%	1.29%	9.23%	99.80%	**17.79%**	**1.87**
24				Unbounded Steps	8.48%	1.58%	9.09%	98.42%	**17.75%**	**1.84**

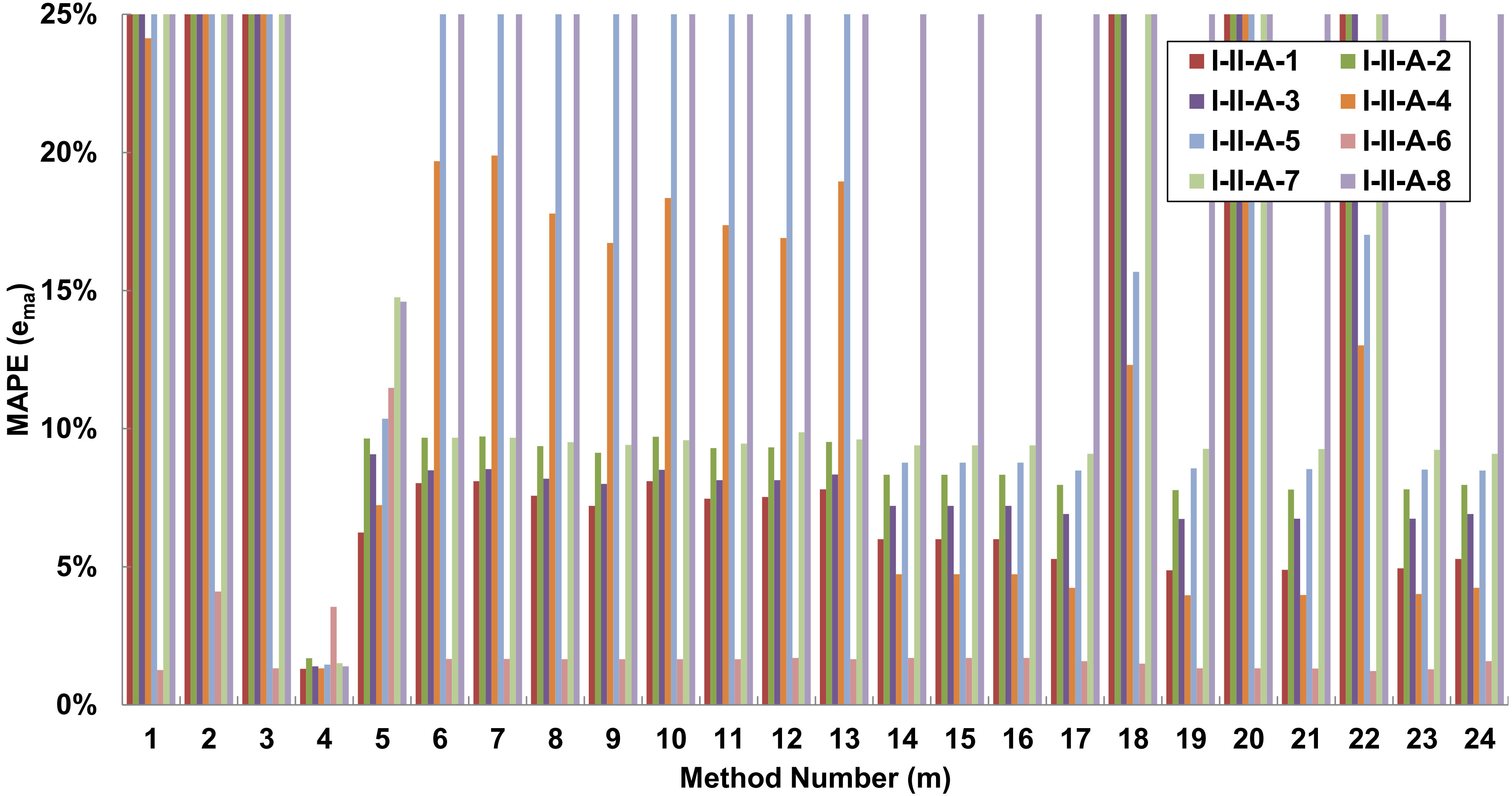


Figure H.6: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-6*

Table H.7: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-7*

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1 *I-II-A-1*	Test Set T_o, o=2 *I-II-A-2*	Test Set T_o, o=3 *I-II-A-3*	Test Set T_o, o=4 *I-II-A-4*
1	Decision Tree	Boosting			91.69%	46.97%	88.81%	46.97%
2		Bootstrap Aggregating			92.79%	91.11%	90.29%	53.97%
3	Neural Network	Feedforward			96.46%	95.59%	95.20%	53.42%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.59%	8.56%	7.62%	6.23%
7			Andrews		5.57%	8.55%	7.61%	6.21%
8			Cauchy M-estimators by Moore		6.95%	9.01%	8.01%	7.13%
9			Fair by Rey		6.95%	9.34%	8.26%	7.78%
10			Huber		6.58%	9.11%	8.08%	7.29%
11			Logistic Regression		6.64%	9.15%	8.11%	7.39%
12			Hinch and Talwar		6.52%	9.08%	8.06%	7.59%
13			Holland and Welsch		5.93%	8.74%	7.78%	6.58%
14		Regularisation	Lasso		8.79%	10.48%	9.15%	6.67%
15			Ridge Regression		8.79%	10.48%	9.15%	6.67%
16			Elastic Nets		8.79%	10.48%	9.15%	6.67%
17		Stepwise	Interaction	Bounded Steps	5.82%	4.44%	3.68%	5.47%
18				Unbounded Steps	10.40%	9.97%	10.69%	7.78%
19			Purequadratic	Bounded Steps	8.73%	10.44%	9.11%	6.64%
20				Unbounded Steps	9.68%	9.21%	9.84%	7.28%
21			Quadratic	Bounded Steps	5.82%	4.44%	3.68%	5.47%
22				Unbounded Steps	9.68%	9.21%	9.84%	7.28%
23			Polynomial	Bounded Steps	5.82%	4.44%	3.68%	5.47%
24				Unbounded Steps	9.67%	9.20%	9.83%	7.28%

Table H.7: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-7* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=5 *I-II-A-5*	Test Set T_o, o=6 *I-II-A-6*	Test Set T_u, u=1 *I-II-A-7*	Test Set T_o, o=7 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			8.34%	25.68%	1.96%	384.90%	**86.92%**	**1.44**
2	Decision Tree	Bootstrap Aggregating			13.50%	35.69%	4.34%	397.38%	**97.38%**	**1.30**
3	Neural Network	Feedforward			8.08%	24.92%	1.98%	383.66%	**94.91%**	**1.30**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		9.09%	4.62%	2.42%	56.30%	**12.55%**	**1.42**
7	Multiple Linear Regression	Robust	Andrews		9.11%	4.63%	2.42%	56.36%	**12.56%**	**1.42**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		9.28%	4.72%	2.41%	55.07%	**12.82%**	**1.34**
9	Multiple Linear Regression	Robust	Fair by Rey		9.74%	4.86%	2.41%	55.28%	**13.08%**	**1.32**
10	Multiple Linear Regression	Robust	Huber		9.30%	4.74%	2.42%	55.31%	**12.85%**	**1.35**
11	Multiple Linear Regression	Robust	Logistic Regression		9.44%	4.78%	2.41%	55.16%	**12.89%**	**1.34**
12	Multiple Linear Regression	Robust	Hinch and Talwar		10.29%	4.69%	2.42%	55.71%	**13.05%**	**1.34**
13	Multiple Linear Regression	Robust	Holland and Welsch		9.04%	4.60%	2.42%	55.35%	**12.56%**	**1.39**
14	Multiple Linear Regression	Regularisation	Lasso		4.08%	5.29%	2.43%	57.28%	**13.02%**	**1.39**
15	Multiple Linear Regression	Regularisation	Ridge Regression		4.08%	5.29%	2.43%	57.28%	**13.02%**	**1.39**
16	Multiple Linear Regression	Regularisation	Elastic Nets		4.08%	5.29%	2.43%	57.28%	**13.02%**	**1.39**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	4.03%	2.02%	2.43%	12.69%	**5.07%**	**0.66**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	4.17%	3.81%	2.08%	41.18%	**11.26%**	**1.11**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	4.07%	5.26%	2.43%	57.28%	**13.00%**	**1.39**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	3.94%	3.37%	1.91%	38.96%	**10.52%**	**1.13**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	4.03%	2.02%	2.43%	12.69%	**5.07%**	**0.66**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	3.94%	3.37%	1.91%	38.96%	**10.52%**	**1.13**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	4.03%	2.02%	2.43%	12.69%	**5.07%**	**0.66**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	3.95%	3.37%	1.92%	38.88%	**10.51%**	**1.13**

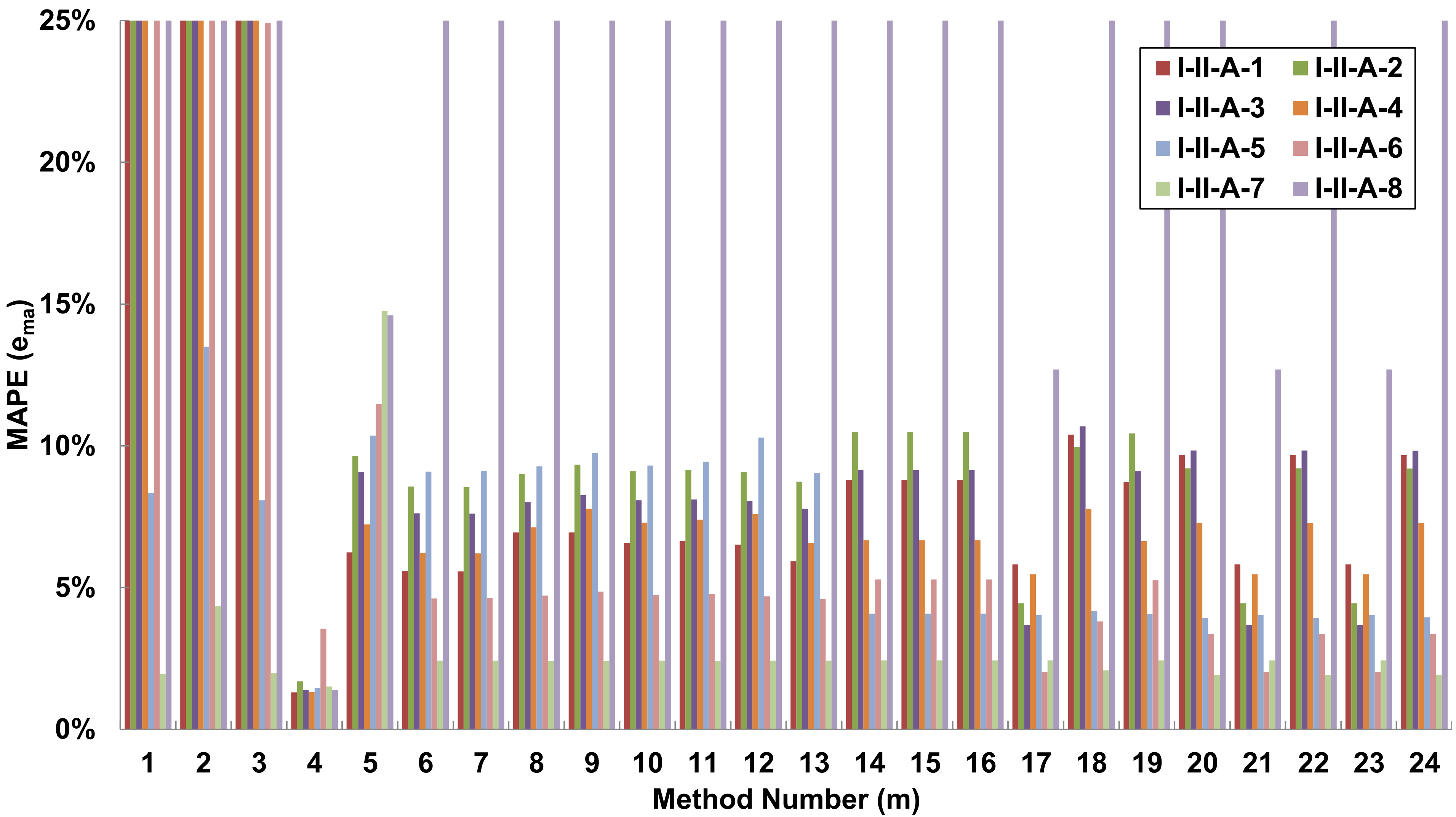


Figure H.7: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-7*

Table H.8: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8*

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1 *I-II-A-1*	Test Set T_o, o=2 *I-II-A-2*	Test Set T_o, o=3 *I-II-A-3*	Test Set T_o, o=4 *I-II-A-4*
1	Decision Tree	Boosting			98.21%	88.12%	97.59%	88.57%
2	Decision Tree	Bootstrap Aggregating			98.57%	98.24%	98.08%	90.89%
3	Neural Network	Feedforward			98.43%	98.05%	97.88%	89.95%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		6.09%	9.33%	8.44%	6.65%
7	Multiple Linear Regression	Robust	Andrews		6.07%	9.32%	8.43%	6.63%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		7.07%	9.89%	8.94%	7.72%
9	Multiple Linear Regression	Robust	Fair by Rey		8.03%	10.50%	9.39%	8.76%
10	Multiple Linear Regression	Robust	Huber		7.46%	10.13%	9.13%	8.14%
11	Multiple Linear Regression	Robust	Logistic Regression		7.52%	10.16%	9.15%	8.21%
12	Multiple Linear Regression	Robust	Hinch and Talwar		5.97%	9.27%	8.37%	6.44%
13	Multiple Linear Regression	Robust	Holland and Welsch		6.33%	9.45%	8.57%	6.90%
14	Multiple Linear Regression	Regularisation	Lasso		98.64%	98.32%	98.17%	91.33%
15	Multiple Linear Regression	Regularisation	Ridge Regression		98.64%	98.32%	98.17%	91.33%
16	Multiple Linear Regression	Regularisation	Elastic Nets		98.64%	98.32%	98.17%	91.33%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	5.22%	4.10%	3.11%	4.96%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	12.97%	13.10%	14.02%	12.34%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	10.41%	12.07%	10.79%	10.58%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	13.54%	13.69%	14.64%	12.86%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	5.22%	4.10%	3.11%	4.96%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	13.54%	13.69%	14.64%	12.86%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	5.22%	4.10%	3.11%	4.96%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	13.55%	13.70%	14.65%	12.86%

Table H.8: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8* (*cont.*)

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma}: Test Set $T_o, o=5$ *I-II-A-5*	Test Set $T_o, o=6$ *I-II-A-6*	Test Set $T_o, o=7$ *I-II-A-7*	Test Set $T_u, u=1$ *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			77.78%	84.00%	73.43%	1.87%	**76.20%**	**0.41**
2	Decision Tree	Bootstrap Aggregating			82.31%	87.30%	78.82%	5.24%	**79.93%**	**0.39**
3	Neural Network	Feedforward			81.59%	86.11%	78.42%	6.48%	**79.61%**	**0.38**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		8.26%	9.68%	11.34%	2.30%	**7.76%**	**0.36**
7	Multiple Linear Regression	Robust	Andrews		8.25%	9.67%	11.35%	2.30%	**7.75%**	**0.36**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		8.96%	10.14%	11.51%	2.29%	**8.32%**	**0.34**
9	Multiple Linear Regression	Robust	Fair by Rey		9.63%	10.58%	11.77%	2.29%	**8.87%**	**0.33**
10	Multiple Linear Regression	Robust	Huber		9.23%	10.30%	11.60%	2.30%	**8.54%**	**0.33**
11	Multiple Linear Regression	Robust	Logistic Regression		9.28%	10.33%	11.62%	2.29%	**8.57%**	**0.33**
12	Multiple Linear Regression	Robust	Hinch and Talwar		8.07%	9.63%	11.33%	2.30%	**7.67%**	**0.36**
13	Multiple Linear Regression	Robust	Holland and Welsch		8.42%	9.78%	11.36%	2.29%	**7.89%**	**0.35**
14	Multiple Linear Regression	Regularisation	Lasso		83.15%	87.81%	79.89%	3.51%	**80.10%**	**0.40**
15	Multiple Linear Regression	Regularisation	Ridge Regression		83.15%	87.81%	79.89%	3.51%	**80.10%**	**0.40**
16	Multiple Linear Regression	Regularisation	Elastic Nets		83.15%	87.81%	79.89%	3.51%	**80.10%**	**0.40**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	3.44%	2.90%	3.59%	2.30%	**3.70%**	**0.27**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	11.05%	9.62%	10.88%	2.02%	**10.75%**	**0.35**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	10.38%	11.80%	12.48%	2.33%	**10.11%**	**0.32**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	11.52%	10.22%	11.36%	1.91%	**11.22%**	**0.36**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	3.44%	2.90%	3.59%	2.30%	**3.70%**	**0.27**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	11.52%	10.22%	11.36%	1.91%	**11.22%**	**0.36**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	3.44%	2.90%	3.59%	2.30%	**3.70%**	**0.27**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	11.53%	10.23%	11.36%	1.91%	**11.22%**	**0.36**

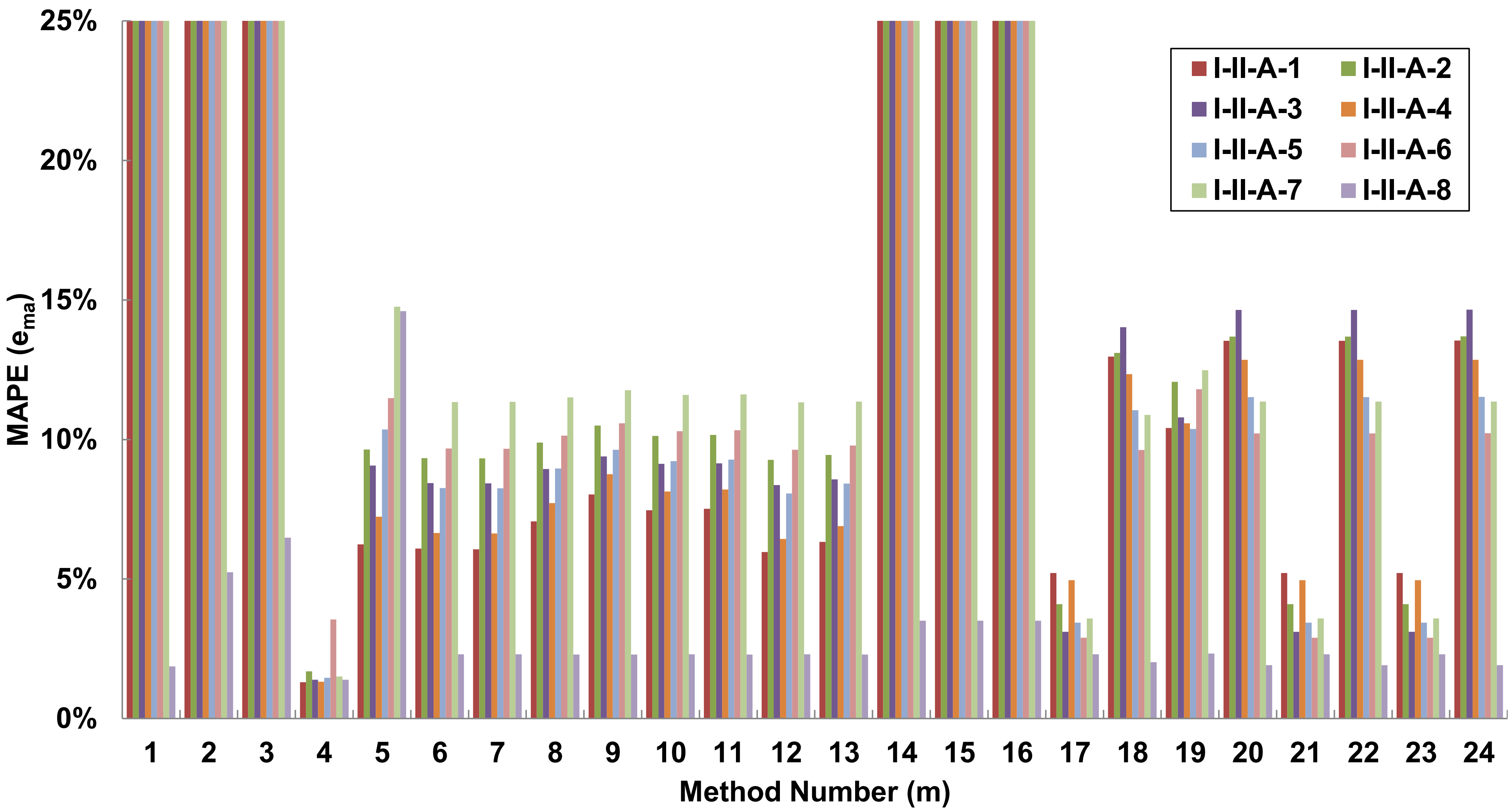


Figure H.8: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8*

Table H.9: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,1*

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_u, u=2$ *I-II-A-1*	e_{ma} Test Set $T_o, o=1$ *I-II-A-2*	e_{ma} Test Set $T_o, o=2$ *I-II-A-3*	e_{ma} Test Set $T_o, o=3$ *I-II-A-4*
1	Decision Tree	Boosting			4.08%	469.19%	14.00%	383.16%
2	Decision Tree	Bootstrap Aggregating			16.26%	19.28%	14.14%	378.93%
3	Neural Network	Feedforward			2.93%	81.94%	134.25%	147.76%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		3.94%	6.49%	5.40%	12.25%
7	Multiple Linear Regression	Robust	Andrews		3.94%	6.49%	5.39%	12.25%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		4.76%	7.89%	6.94%	5.81%
9	Multiple Linear Regression	Robust	Fair by Rey		3.79%	6.36%	5.45%	24.47%
10	Multiple Linear Regression	Robust	Huber		3.78%	6.18%	5.12%	18.27%
11	Multiple Linear Regression	Robust	Logistic Regression		3.77%	6.17%	5.11%	18.29%
12	Multiple Linear Regression	Robust	Hinch and Talwar		5.77%	9.23%	8.34%	6.46%
13	Multiple Linear Regression	Robust	Holland and Welsch		3.94%	6.50%	5.41%	12.20%
14	Multiple Linear Regression	Regularisation	Lasso		4.02%	4.57%	3.28%	30.03%
15	Multiple Linear Regression	Regularisation	Ridge Regression		4.26%	4.51%	2.70%	33.44%
16	Multiple Linear Regression	Regularisation	Elastic Nets		4.26%	4.51%	2.70%	33.44%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	3.79%	4.29%	4.01%	35.52%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	3.80%	4.41%	4.36%	36.41%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	4.10%	4.67%	2.74%	31.83%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	3.38%	7.20%	8.78%	50.90%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	3.41%	4.25%	4.03%	27.94%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	3.43%	4.40%	4.36%	28.91%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	2.96%	3.78%	2.34%	3.82%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	2.85%	3.43%	3.18%	11.42%

Table H.9: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,1* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=4 *I-II-A-5*	Test Set T_o, o=5 *I-II-A-6*	Test Set T_o, o=6 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			878.38%	121.82%	156.20%	414.25%	**305.14%**	**0.96**
2		Bootstrap Aggregating			585.45%	87.32%	77.88%	14.96%	**149.28%**	**1.44**
3	Neural Network	Feedforward			560.14%	18.67%	17.13%	10.74%	**121.70%**	**1.53**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		28.36%	8.12%	27.87%	186.85%	**34.91%**	**1.78**
7			Andrews		28.37%	8.13%	27.88%	186.93%	**34.92%**	**1.78**
8			Cauchy M-estimators by Moore		14.61%	4.62%	10.43%	89.79%	**18.11%**	**1.61**
9			Fair by Rey		64.79%	15.96%	42.80%	261.53%	**53.14%**	**1.63**
10			Huber		46.31%	11.85%	35.67%	224.60%	**43.97%**	**1.70**
11			Logistic Regression		46.49%	11.81%	35.86%	225.62%	**44.14%**	**1.70**
12			Hinch and Talwar		9.47%	10.08%	11.84%	2.29%	**7.94%**	**0.38**
13			Holland and Welsch		28.29%	8.06%	27.67%	185.85%	**34.74%**	**1.78**
14		Regularisation	Lasso		62.96%	36.02%	77.44%	438.76%	**82.14%**	**1.79**
15			Ridge Regression		61.19%	29.71%	70.74%	405.04%	**76.45%**	**1.77**
16			Elastic Nets		61.19%	29.71%	70.74%	405.04%	**76.45%**	**1.77**
17		Stepwise	Interaction	Bounded Steps	70.88%	8.42%	7.93%	2.54%	**17.17%**	**1.41**
18				Unbounded Steps	69.86%	8.40%	8.02%	4.43%	**17.46%**	**1.36**
19			Purequadratic	Bounded Steps	56.19%	34.01%	75.05%	425.30%	**79.24%**	**1.80**
20				Unbounded Steps	151.42%	32.70%	29.25%	17.17%	**37.60%**	**1.29**
21			Quadratic	Bounded Steps	53.65%	6.33%	6.42%	2.52%	**13.57%**	**1.34**
22				Unbounded Steps	53.48%	5.89%	6.22%	4.42%	**13.89%**	**1.30**
23			Polynomial	Bounded Steps	3.42%	3.96%	4.13%	4.00%	**3.55%**	**0.17**
24				Unbounded Steps	29.15%	8.63%	8.97%	2.94%	**8.82%**	**1.01**

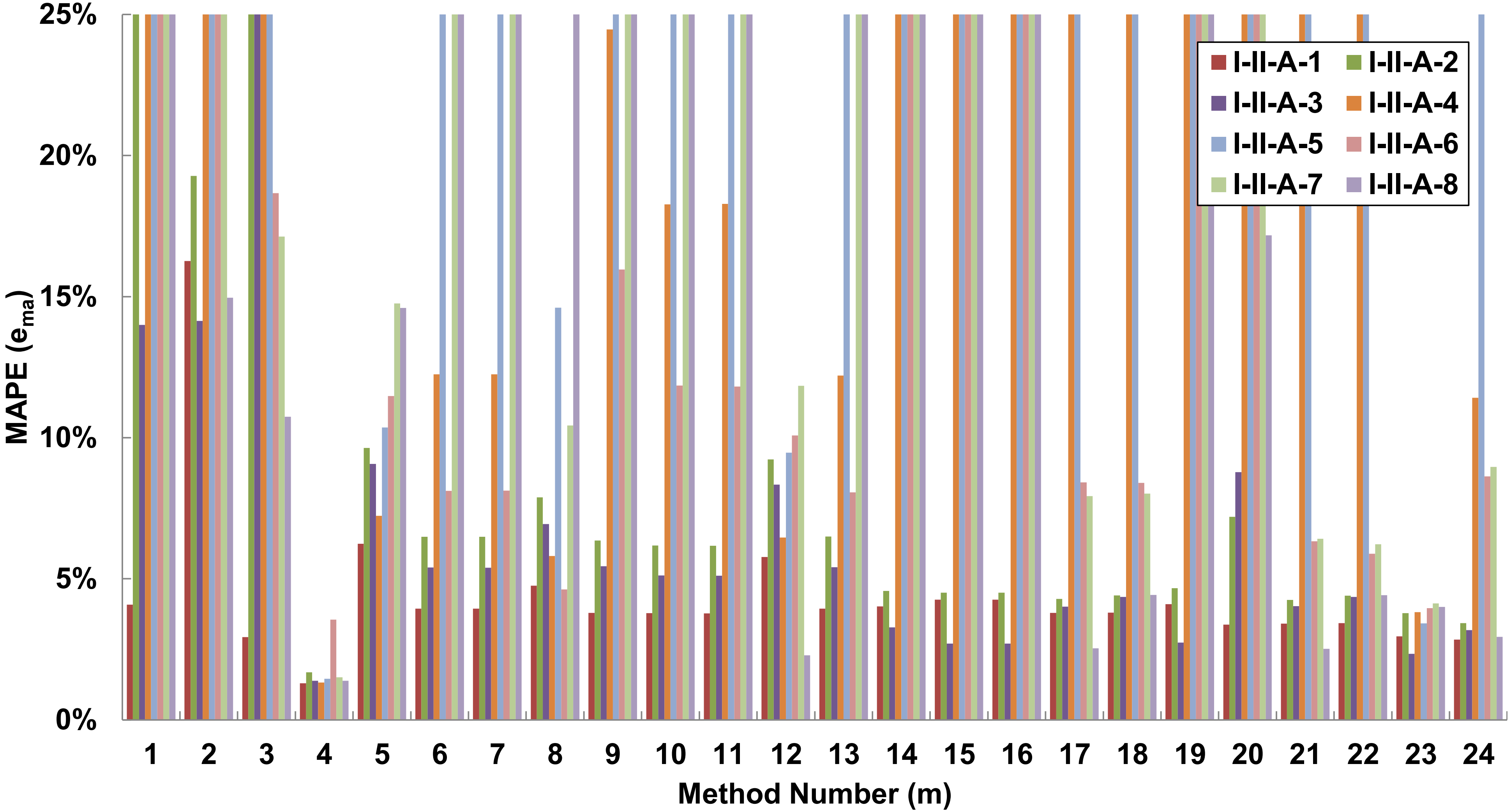


Figure H.9: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,1*

Table H.10: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,2*

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ *I-II-A-1*	Test Set $T_u, u=2$ *I-II-A-2*	Test Set $T_o, o=2$ *I-II-A-3*	Test Set $T_o, o=3$ *I-II-A-4*
1	Decision Tree	Boosting			7.19%	5.45%	4.58%	345.95%
2		Bootstrap Aggregating						
3	Neural Network	Feedforward						
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		3.66%	4.05%	2.56%	32.48%
7			Andrews		3.66%	4.05%	2.56%	32.48%
8			Cauchy M-estimators by Moore		3.80%	3.77%	2.36%	35.88%
9			Fair by Rey		4.08%	3.80%	2.57%	39.04%
10			Huber		3.98%	3.77%	2.51%	37.84%
11			Logistic Regression		3.99%	3.77%	2.52%	37.94%
12			Hinch and Talwar		3.69%	4.06%	2.57%	32.78%
13			Holland and Welsch		3.65%	4.02%	2.52%	32.59%
14		Regularisation	Lasso					
15			Ridge Regression					
16			Elastic Nets					
17		Stepwise	Interaction	Bounded Steps	3.57%	3.32%	2.31%	21.54%
18				Unbounded Steps	3.46%	3.32%	2.32%	21.77%
19			Purequadratic	Bounded Steps	6.37%	3.96%	3.06%	61.63%
20				Unbounded Steps	3.97%	3.33%	2.29%	21.03%
21			Quadratic	Bounded Steps	3.88%	3.33%	2.19%	22.72%
22				Unbounded Steps	4.09%	3.33%	2.20%	22.87%
23			Polynomial	Bounded Steps	4.12%	3.28%	2.22%	20.41%
24				Unbounded Steps				

Table H.10: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,2* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=4$ *I-II-A-5*	Test Set $T_o, o=5$ *I-II-A-6*	Test Set $T_o, o=6$ *I-II-A-7*	Test Set $T_u, u=1$ *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			749.22%	83.12%	71.65%	45.61%	**164.10%**	**1.60**
2	Decision Tree	Bootstrap Aggregating								
3	Neural Network	Feedforward								
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		63.59%	29.23%	82.02%	458.65%	**84.53%**	**1.82**
7	Multiple Linear Regression	Robust	Andrews		63.56%	29.22%	82.01%	458.58%	**84.52%**	**1.82**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		74.07%	37.26%	93.83%	518.67%	**96.21%**	**1.81**
9	Multiple Linear Regression	Robust	Fair by Rey		80.02%	43.10%	101.13%	555.47%	**103.65%**	**1.80**
10	Multiple Linear Regression	Robust	Huber		78.05%	40.93%	98.97%	544.25%	**101.29%**	**1.80**
11	Multiple Linear Regression	Robust	Logistic Regression		78.22%	41.24%	99.19%	545.44%	**101.54%**	**1.80**
12	Multiple Linear Regression	Robust	Hinch and Talwar		64.01%	29.48%	82.74%	462.12%	**85.18%**	**1.82**
13	Multiple Linear Regression	Robust	Holland and Welsch		64.09%	29.52%	82.43%	460.95%	**84.97%**	**1.82**
14	Multiple Linear Regression	Regularisation	Lasso							
15	Multiple Linear Regression	Regularisation	Ridge Regression							
16	Multiple Linear Regression	Regularisation	Elastic Nets							
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	16.01%	4.14%	7.40%	35.86%	**11.77%**	**1.02**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	16.06%	4.48%	7.73%	35.86%	**11.88%**	**1.01**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	91.89%	57.61%	118.45%	641.06%	**123.00%**	**1.74**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	12.29%	4.29%	7.52%	37.31%	**11.50%**	**1.06**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	12.79%	7.47%	14.02%	77.16%	**17.95%**	**1.39**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	12.06%	6.06%	7.81%	37.32%	**11.97%**	**1.02**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	7.78%	4.32%	7.35%	36.75%	**10.78%**	**1.11**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps						

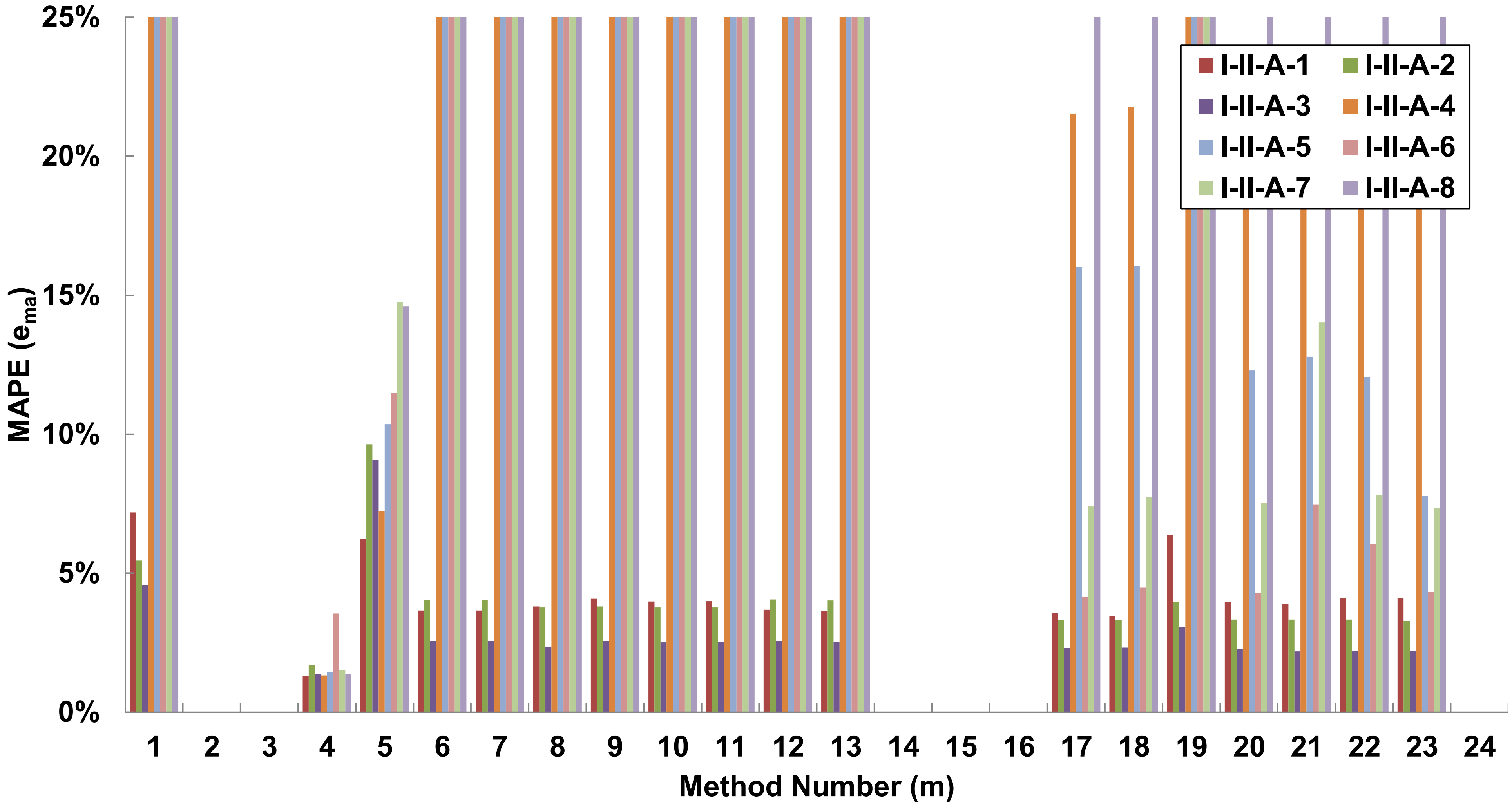


Figure H.10: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,2*

Table H.11: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3*

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ *I-II-A-1*	e_{ma} Test Set $T_o, o=2$ *I-II-A-2*	e_{ma} Test Set $T_u, u=2$ *I-II-A-3*	e_{ma} Test Set $T_o, o=3$ *I-II-A-4*
1	Decision Tree	Boosting			13.94%	460.54%	6.69%	354.64%
2		Bootstrap Aggregating			22.56%	11.90%	5.92%	287.79%
3	Neural Network	Feedforward			76.10%	23.21%	1.91%	453.70%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		13.40%	13.07%	11.19%	13.02%
7			Andrews		13.40%	13.07%	11.19%	13.02%
8			Cauchy M-estimators by Moore		24.65%	12.22%	7.54%	108.98%
9			Fair by Rey		24.43%	12.20%	7.51%	107.67%
10			Huber		24.52%	12.21%	7.52%	108.23%
11			Logistic Regression		24.52%	12.21%	7.52%	108.22%
12			Hinch and Talwar		9.05%	9.70%	8.10%	31.48%
13			Holland and Welsch		6.61%	9.33%	8.26%	13.38%
14		Regularisation	Lasso		8.84%	9.59%	7.37%	72.05%
15			Ridge Regression		17.25%	10.81%	7.37%	48.18%
16			Elastic Nets		17.25%	10.81%	7.37%	48.18%
17		Stepwise	Interaction	Bounded Steps	3.82%	3.48%	2.09%	3.74%
18				Unbounded Steps	4.43%	3.87%	2.06%	31.41%
19			Purequadratic	Bounded Steps	58.78%	15.90%	7.01%	532.76%
20				Unbounded Steps	7.86%	4.12%	2.01%	55.44%
21			Quadratic	Bounded Steps	3.82%	3.48%	2.09%	3.74%
22				Unbounded Steps	10.18%	4.71%	2.01%	72.04%
23			Polynomial	Bounded Steps	3.82%	3.48%	2.09%	3.74%
24				Unbounded Steps	297.86%	46.02%	1.86%	673.67%

Table H.11: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	Test Set T_o, o=4 *I-II-A-5*	Test Set T_o, o=5 *I-II-A-6*	Test Set T_o, o=6 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
	Method				e_{ma}					
1	Decision Tree	Boosting			780.69%	91.78%	90.41%	74.17%	**234.11%**	**1.18**
2	Decision Tree	Bootstrap Aggregating			646.07%	87.04%	78.42%	4.94%	**143.08%**	**1.56**
3	Neural Network	Feedforward			172.21%	19.35%	20.94%	2.19%	**96.20%**	**1.61**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		32.51%	13.74%	13.79%	2.32%	**14.13%**	**0.59**
7	Multiple Linear Regression	Robust	Andrews		32.51%	13.74%	13.79%	2.32%	**14.13%**	**0.59**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		326.22%	28.89%	24.89%	2.48%	**66.98%**	**1.64**
9	Multiple Linear Regression	Robust	Fair by Rey		322.26%	28.62%	24.66%	2.76%	**66.26%**	**1.64**
10	Multiple Linear Regression	Robust	Huber		323.95%	28.73%	24.75%	2.59%	**66.56%**	**1.64**
11	Multiple Linear Regression	Robust	Logistic Regression		323.91%	28.73%	24.76%	2.59%	**66.56%**	**1.64**
12	Multiple Linear Regression	Robust	Hinch and Talwar		88.57%	11.83%	12.50%	2.31%	**21.69%**	**1.31**
13	Multiple Linear Regression	Robust	Holland and Welsch		32.60%	10.06%	11.46%	2.30%	**11.75%**	**0.77**
14	Multiple Linear Regression	Regularisation	Lasso		146.96%	12.81%	13.37%	82.61%	**44.20%**	**1.16**
15	Multiple Linear Regression	Regularisation	Ridge Regression		146.96%	12.81%	13.37%	82.61%	**42.42%**	**1.17**
16	Multiple Linear Regression	Regularisation	Elastic Nets		146.96%	12.81%	13.37%	82.61%	**42.42%**	**1.17**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	3.21%	4.02%	4.57%	2.60%	**3.44%**	**0.23**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	51.84%	4.42%	4.60%	2.29%	**13.12%**	**1.41**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	866.48%	148.95%	144.01%	76.07%	**231.25%**	**1.33**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	123.71%	20.65%	19.66%	7.35%	**30.10%**	**1.38**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	3.21%	4.02%	4.57%	2.60%	**3.44%**	**0.23**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	123.71%	20.65%	19.66%	7.35%	**32.54%**	**1.33**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	3.21%	4.02%	4.57%	2.60%	**3.44%**	**0.23**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	1323.30%	103.07%	94.76%	5.72%	**318.28%**	**1.46**

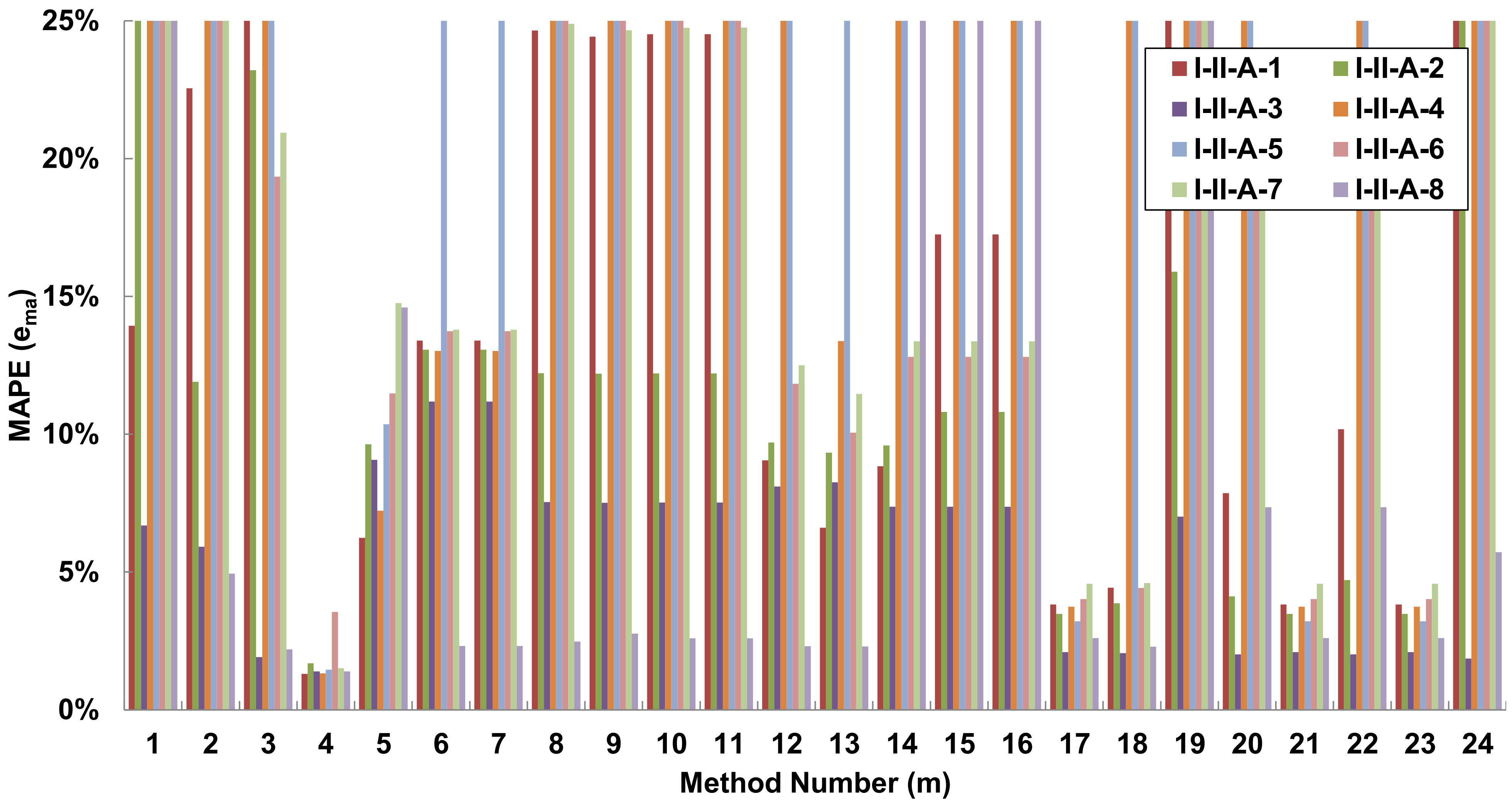


Figure H.11: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3*

Table H.12: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,4*

Number (m)	Method				e_{ma}			
					Test Set $T_o, o=1$	Test Set $T_o, o=2$	Test Set $T_o, o=3$	Test Set $T_u, u=2$
	Class I	Class II	Class III	Class IV	*I-II-A-1*	*I-II-A-2*	*I-II-A-3*	*I-II-A-4*
1	Decision Tree	Boosting			6.24%	42.37%	7.52%	4.48%
2		Bootstrap Aggregating			35.32%	39.82%	39.48%	16.56%
3	Neural Network	Feedforward			4.48%	4.82%	4.53%	2.31%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		6.06%	9.40%	8.77%	6.04%
7			Andrews		6.06%	9.40%	8.77%	6.04%
8			Cauchy M-estimators by Moore		6.05%	9.39%	8.76%	6.01%
9			Fair by Rey		5.73%	9.01%	8.10%	5.03%
10			Huber		5.71%	9.13%	8.24%	5.44%
11			Logistic Regression		5.70%	9.12%	8.21%	5.41%
12			Hinch and Talwar		6.05%	9.40%	8.77%	6.03%
13			Holland and Welsch		6.06%	9.40%	8.77%	6.04%
14		Regularisation	Lasso		5.82%	8.79%	7.84%	4.20%
15			Ridge Regression		5.82%	8.83%	7.87%	4.08%
16			Elastic Nets		5.82%	8.79%	7.84%	4.20%
17		Stepwise	Interaction	Bounded Steps	4.03%	7.42%	6.74%	3.78%
18				Unbounded Steps	4.13%	7.21%	6.25%	2.81%
19			Purequadratic	Bounded Steps	5.87%	8.91%	7.97%	4.71%
20				Unbounded Steps	4.20%	6.94%	5.86%	2.76%
21			Quadratic	Bounded Steps	4.24%	6.78%	5.60%	2.77%
22				Unbounded Steps	4.11%	7.12%	6.15%	2.83%
23			Polynomial	Bounded Steps	4.73%	4.61%	3.80%	2.51%
24				Unbounded Steps	20.40%	22.98%	115.02%	2.17%

Table H.12: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,4* (*cont.*)

	Method				$\mathbf{e_{ma}}$					
					Test Set T_o,o=4	Test Set T_o,o=5	Test Set T_o,o=6	Test Set T_u,u=1		
Number (m)	**Class I**	**Class II**	**Class III**	**Class IV**	***I-II-A-5***	***I-II-A-6***	***I-II-A-7***	***I-II-A-8***	μ_e	c_e
1	Decision Tree	Boosting			43.23%	87.12%	83.68%	39.18%	**39.23%**	**0.84**
2	Decision Tree	Bootstrap Aggregating			54.97%	77.39%	62.60%	5.21%	**41.42%**	**0.57**
3	Neural Network	Feedforward			3.74%	22.02%	21.75%	2.25%	**8.24%**	**1.03**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.89%	10.12%	11.89%	2.33%	**7.81%**	**0.38**
7	Multiple Linear Regression	Robust	Andrews		7.89%	10.13%	11.89%	2.33%	**7.81%**	**0.38**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		7.84%	10.09%	11.84%	2.33%	**7.79%**	**0.38**
9	Multiple Linear Regression	Robust	Fair by Rey		5.80%	8.05%	8.30%	18.93%	**8.62%**	**0.51**
10	Multiple Linear Regression	Robust	Huber		6.76%	9.47%	10.53%	8.73%	**8.00%**	**0.23**
11	Multiple Linear Regression	Robust	Logistic Regression		6.73%	9.41%	10.42%	9.09%	**8.01%**	**0.23**
12	Multiple Linear Regression	Robust	Hinch and Talwar		7.86%	10.11%	11.87%	2.31%	**7.80%**	**0.38**
13	Multiple Linear Regression	Robust	Holland and Welsch		7.89%	10.12%	11.88%	2.32%	**7.81%**	**0.38**
14	Multiple Linear Regression	Regularisation	Lasso		4.12%	6.31%	5.37%	41.71%	**10.52%**	**1.21**
15	Multiple Linear Regression	Regularisation	Ridge Regression		4.01%	6.10%	4.21%	40.31%	**10.15%**	**1.21**
16	Multiple Linear Regression	Regularisation	Elastic Nets		4.12%	6.31%	5.37%	41.71%	**10.52%**	**1.21**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	3.98%	34.13%	30.04%	3.08%	**11.65%**	**1.09**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	5.89%	10.97%	11.10%	3.65%	**6.50%**	**0.48**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	4.99%	7.17%	7.15%	43.78%	**11.32%**	**1.17**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	5.63%	8.92%	9.82%	3.61%	**5.97%**	**0.42**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	6.08%	8.53%	9.60%	2.56%	**5.77%**	**0.44**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	5.81%	10.77%	10.94%	3.70%	**6.43%**	**0.48**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	3.23%	44.39%	53.07%	3.21%	**14.94%**	**1.40**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	10.64%	691.25%	468.35%	2.55%	**166.67%**	**1.59**

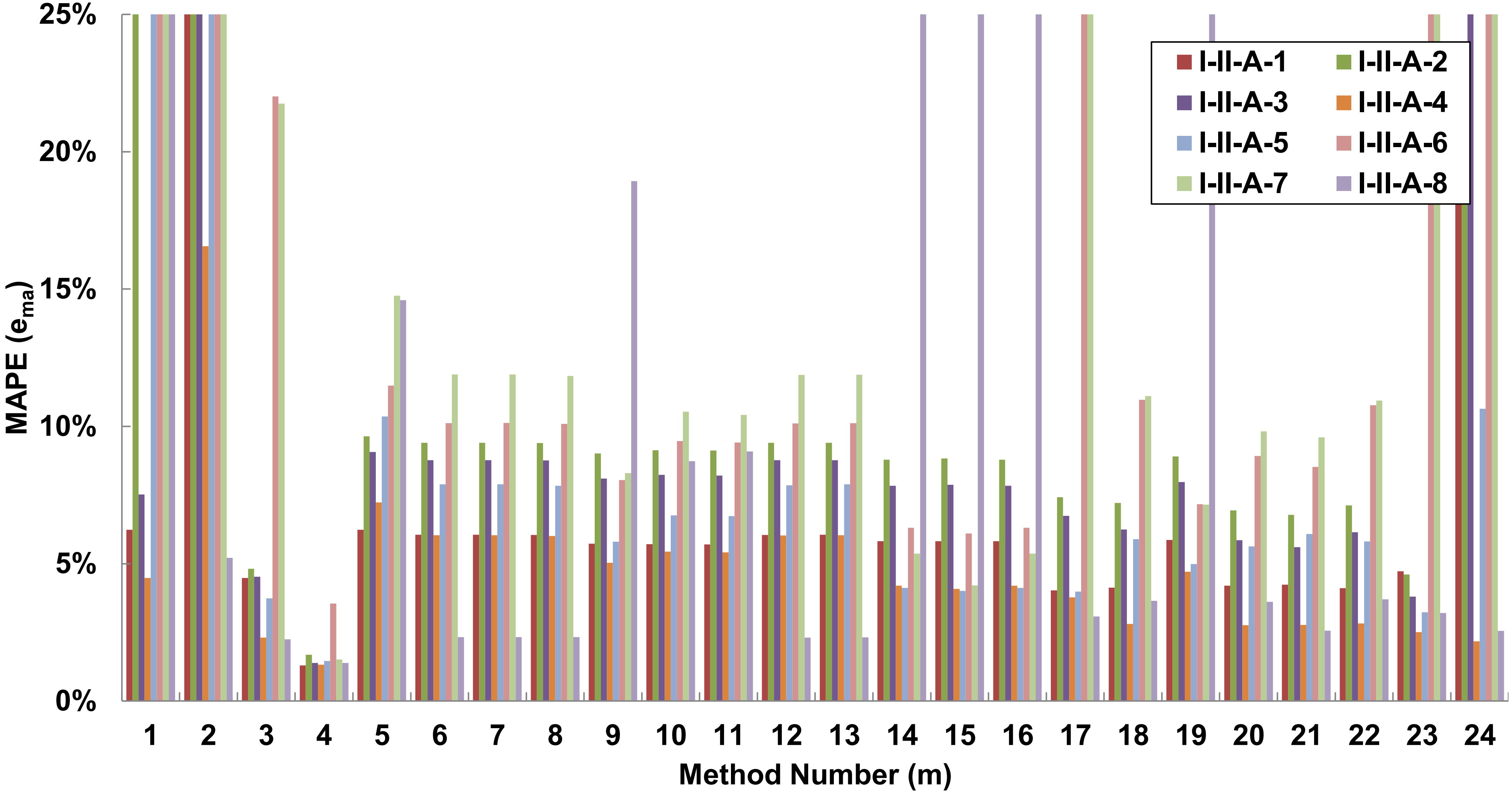


Figure H.12: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,4*

Table H.13: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,5*

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1 *I-II-A-1*	Test Set T_o, o=2 *I-II-A-2*	Test Set T_o, o=3 *I-II-A-3*	Test Set T_o, o=4 *I-II-A-4*
1	Decision Tree	Boosting			66.67%	55.56%	55.05%	11.39%
2		Bootstrap Aggregating			86.15%	82.91%	81.29%	27.05%
3	Neural Network	Feedforward			72.14%	32.77%	15.13%	55.78%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.93%	9.36%	8.74%	6.14%
7			Andrews		5.93%	9.36%	8.74%	6.14%
8			Cauchy M-estimators by Moore		5.98%	9.37%	8.75%	6.02%
9			Fair by Rey		5.64%	8.92%	7.98%	6.17%
10			Huber		5.63%	9.00%	8.24%	6.23%
11			Logistic Regression		5.62%	8.99%	8.22%	6.24%
12			Hinch and Talwar		5.92%	9.35%	8.72%	6.13%
13			Holland and Welsch		5.94%	9.37%	8.75%	6.13%
14		Regularisation	Lasso		7.43%	9.86%	8.75%	6.39%
15			Ridge Regression		7.42%	9.86%	8.76%	6.39%
16			Elastic Nets		7.43%	9.86%	8.75%	6.39%
17		Stepwise	Interaction	Bounded Steps	3.47%	3.88%	2.66%	3.49%
18				Unbounded Steps	3.47%	3.87%	2.66%	3.63%
19			Purequadratic	Bounded Steps	32.34%	24.65%	20.95%	7.27%
20				Unbounded Steps	3.64%	3.90%	2.58%	3.67%
21			Quadratic	Bounded Steps	3.49%	3.86%	2.63%	3.46%
22				Unbounded Steps	3.49%	3.85%	2.63%	3.60%
23			Polynomial	Bounded Steps	3.56%	3.92%	2.74%	4.21%
24				Unbounded Steps	36.88%	25.58%	22.13%	6.10%

Table H.13: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,5* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_u, u=2 *I-II-A-5*	Test Set T_o, o=5 *I-II-A-6*	Test Set T_o, o=6 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			4.82%	51.62%	86.86%	26.70%	**44.83%**	**0.63**
2		Bootstrap Aggregating			9.50%	63.95%	49.66%	4.62%	**50.64%**	**0.66**
3	Neural Network	Feedforward			2.80%	20.05%	23.93%	2.34%	**28.12%**	**0.88**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.92%	10.05%	11.82%	2.31%	**7.78%**	**0.38**
7			Andrews		7.92%	10.05%	11.82%	2.31%	**7.78%**	**0.38**
8			Cauchy M-estimators by Moore		7.80%	10.04%	11.77%	2.38%	**7.76%**	**0.38**
9			Fair by Rey		5.30%	8.87%	9.12%	21.15%	**9.14%**	**0.56**
10			Huber		5.75%	9.40%	9.91%	16.68%	**8.86%**	**0.40**
11			Logistic Regression		5.67%	9.28%	9.69%	17.14%	**8.86%**	**0.42**
12			Hinch and Talwar		7.91%	10.03%	11.80%	2.31%	**7.77%**	**0.38**
13			Holland and Welsch		7.91%	10.05%	11.82%	2.31%	**7.79%**	**0.38**
14		Regularisation	Lasso		5.02%	7.54%	6.10%	28.17%	**9.91%**	**0.76**
15			Ridge Regression		5.06%	7.66%	6.19%	27.70%	**9.88%**	**0.74**
16			Elastic Nets		5.02%	7.54%	6.10%	28.17%	**9.91%**	**0.76**
17		Stepwise	Linear with Interactions of Linear Terms	Bounded Steps	2.97%	5.55%	5.68%	2.93%	**3.83%**	**0.31**
18				Unbounded Steps	2.97%	5.55%	5.68%	2.93%	**3.85%**	**0.30**
19			Linear and Squared Terms	Bounded Steps	5.20%	7.60%	6.32%	28.77%	**16.64%**	**0.68**
20				Unbounded Steps	2.91%	5.10%	5.68%	3.35%	**3.85%**	**0.27**
21			Linear and Squared Terms including Interactions	Bounded Steps	2.90%	5.52%	5.67%	3.18%	**3.84%**	**0.30**
22				Unbounded Steps	2.90%	5.52%	5.67%	3.18%	**3.86%**	**0.30**
23			Polynomial	Bounded Steps	2.84%	5.52%	5.75%	1.98%	**3.82%**	**0.35**
24				Unbounded Steps	2.81%	3.40%	3.93%	2.10%	**12.87%**	**1.04**

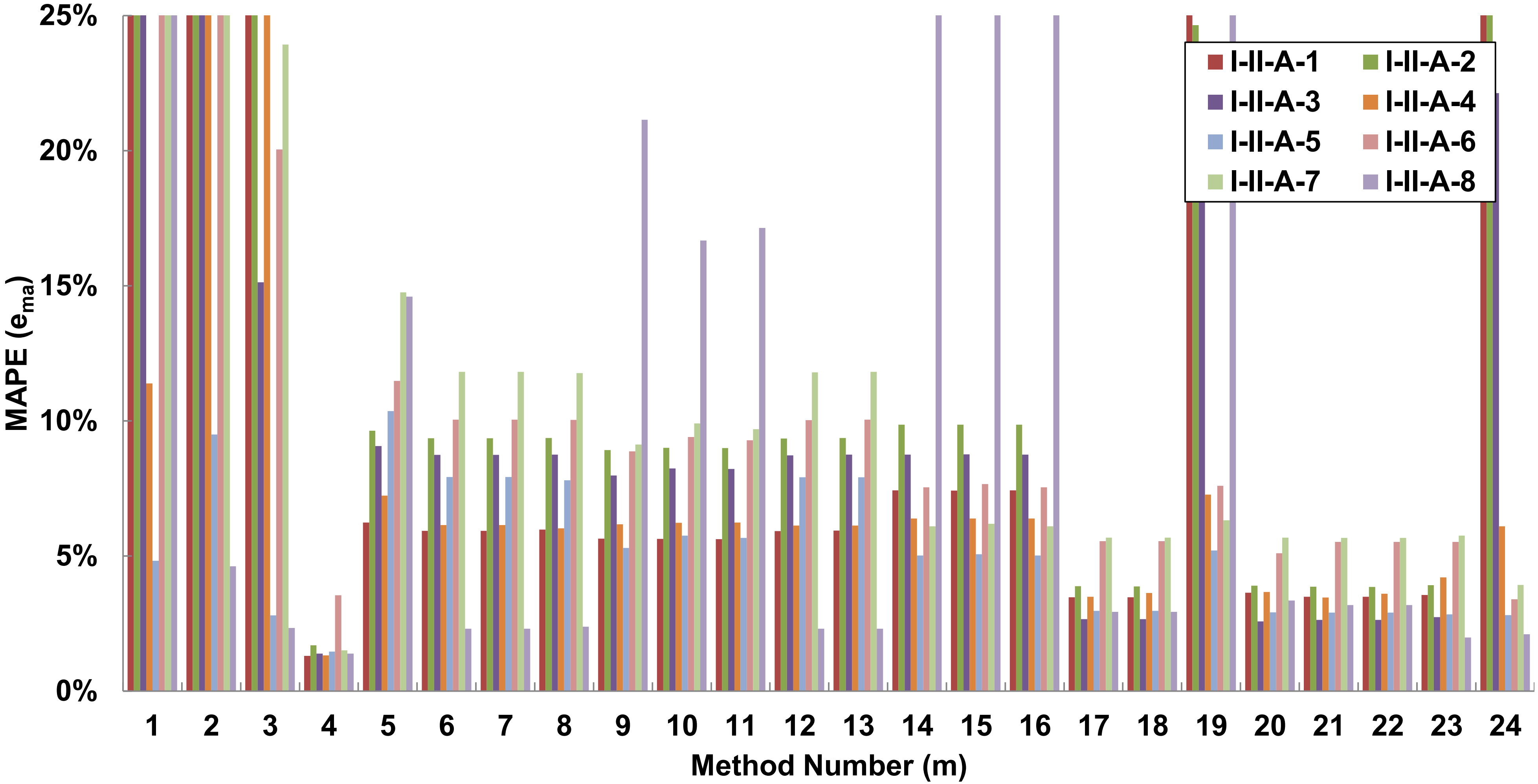


Figure H.13: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,5*

Table H.14: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,6*

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1 *I-II-A-1*	Test Set T_o, o=2 *I-II-A-2*	Test Set T_o, o=3 *I-II-A-3*	Test Set T_o, o=4 *I-II-A-4*
1	Decision Tree	Boosting			87.26%	56.00%	82.83%	23.98%
2	Decision Tree	Bootstrap Aggregating			92.22%	90.41%	89.50%	50.54%
3	Neural Network	Feedforward			85.21%	81.83%	80.19%	22.75%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.77%	9.23%	8.32%	6.23%
7	Multiple Linear Regression	Robust	Andrews		5.77%	9.23%	8.32%	6.24%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		5.77%	9.23%	8.32%	6.15%
9	Multiple Linear Regression	Robust	Fair by Rey		5.92%	9.24%	8.32%	6.10%
10	Multiple Linear Regression	Robust	Huber		5.81%	9.21%	8.30%	6.06%
11	Multiple Linear Regression	Robust	Logistic Regression		5.82%	9.21%	8.30%	6.07%
12	Multiple Linear Regression	Robust	Hinch and Talwar		5.77%	9.23%	8.32%	6.22%
13	Multiple Linear Regression	Robust	Holland and Welsch		5.77%	9.23%	8.32%	6.20%
14	Multiple Linear Regression	Regularisation	Lasso		8.56%	10.72%	9.48%	8.18%
15	Multiple Linear Regression	Regularisation	Ridge Regression		8.56%	10.72%	9.48%	8.18%
16	Multiple Linear Regression	Regularisation	Elastic Nets		8.56%	10.72%	9.48%	8.18%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	6.70%	4.78%	4.85%	6.01%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	6.70%	4.78%	4.85%	6.01%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	489.89%	333.87%	284.90%	47.83%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	6.72%	4.79%	4.86%	5.99%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	6.71%	4.79%	4.85%	5.98%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	6.72%	4.79%	4.86%	5.99%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	6.32%	4.50%	4.47%	5.73%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	6.47%	4.57%	4.65%	5.78%

Table H.14: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,6* (*cont.*)

	Method				e_{ma}					
					Test Set T_o, o=5	Test Set T_u, u=2	Test Set T_o, o=6	Test Set T_u, u=1		
Number (m)	Class I	Class II	Class III	Class IV	*I-II-A-5*	*I-II-A-6*	*I-II-A-7*	*I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			42.93%	8.52%	68.76%	13.81%	**48.01%**	**0.64**
2	Decision Tree	Bootstrap Aggregating			13.11%	31.06%	12.79%	55.01%	**54.33%**	**0.62**
3	Neural Network	Feedforward			50.10%	1.26%	62.64%	2.70%	**48.34%**	**0.73**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		8.00%	9.69%	11.52%	2.31%	**7.63%**	**0.37**
7	Multiple Linear Regression	Robust	Andrews		8.00%	9.70%	11.52%	2.31%	**7.64%**	**0.37**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		7.89%	9.64%	11.46%	2.30%	**7.60%**	**0.37**
9	Multiple Linear Regression	Robust	Fair by Rey		7.75%	9.47%	11.08%	2.56%	**7.56%**	**0.35**
10	Multiple Linear Regression	Robust	Huber		7.78%	9.50%	11.21%	2.43%	**7.54%**	**0.36**
11	Multiple Linear Regression	Robust	Logistic Regression		7.77%	9.50%	11.20%	2.43%	**7.54%**	**0.36**
12	Multiple Linear Regression	Robust	Hinch and Talwar		7.98%	9.70%	11.52%	2.31%	**7.63%**	**0.37**
13	Multiple Linear Regression	Robust	Holland and Welsch		7.96%	9.68%	11.50%	2.30%	**7.62%**	**0.37**
14	Multiple Linear Regression	Regularisation	Lasso		7.47%	9.48%	9.19%	13.92%	**9.63%**	**0.21**
15	Multiple Linear Regression	Regularisation	Ridge Regression		7.47%	9.48%	9.19%	13.92%	**9.63%**	**0.21**
16	Multiple Linear Regression	Regularisation	Elastic Nets		7.47%	9.48%	9.19%	13.92%	**9.63%**	**0.21**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	4.19%	1.61%	2.33%	2.04%	**4.06%**	**0.47**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	4.19%	1.61%	2.33%	2.04%	**4.06%**	**0.47**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	15.01%	9.12%	15.89%	14.62%	**151.39%**	**1.25**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	4.14%	1.54%	2.25%	2.20%	**4.06%**	**0.47**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	4.11%	1.51%	2.20%	2.41%	**4.07%**	**0.46**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	4.14%	1.54%	2.25%	2.20%	**4.06%**	**0.47**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	3.87%	1.23%	1.99%	1.89%	**3.75%**	**0.50**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	3.97%	1.50%	2.13%	2.02%	**3.89%**	**0.47**

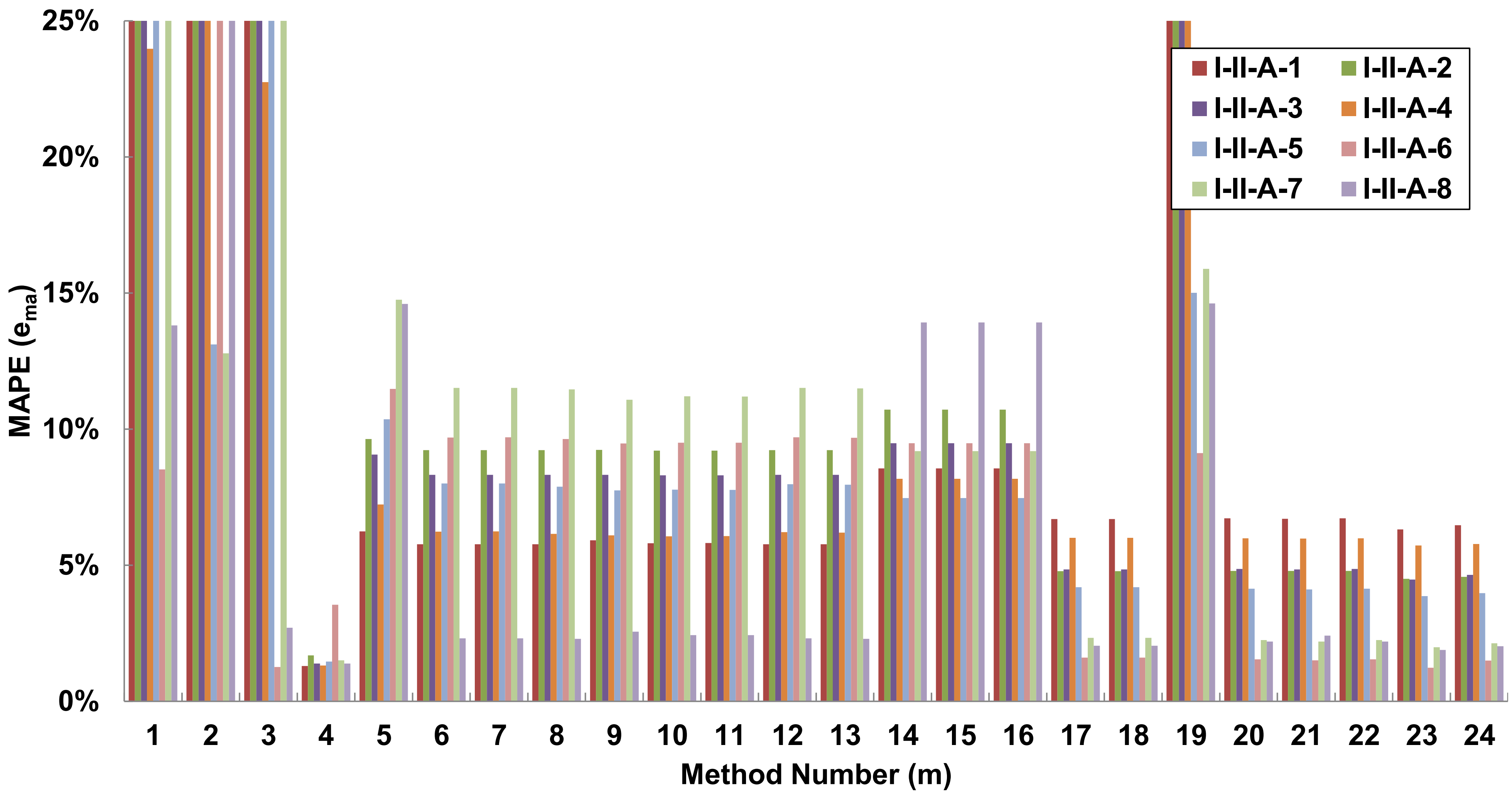


Figure H.14: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,6*

Table H.15: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,7*

	Method				e_{ma}			
					Test Set T_o, o=1	Test Set T_o, o=2	Test Set T_o, o=3	Test Set T_o, o=4
Number (m)	Class I	Class II	Class III	Class IV	*I-II-A-1*	*I-II-A-2*	*I-II-A-3*	*I-II-A-4*
1	Decision Tree	Boosting			91.42%	48.11%	88.44%	45.16%
2		Bootstrap Aggregating			94.05%	92.65%	91.98%	61.98%
3	Neural Network	Feedforward			87.60%	84.24%	82.37%	21.43%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		6.45%	9.27%	8.46%	21.24%
7			Andrews		6.45%	9.27%	8.46%	21.24%
8			Cauchy M-estimators by Moore		6.50%	9.29%	8.48%	21.23%
9			Fair by Rey		8.10%	10.31%	9.15%	10.23%
10			Huber		8.41%	10.51%	9.35%	12.72%
11			Logistic Regression		8.03%	10.27%	9.14%	11.92%
12			Hinch and Talwar		9.71%	11.36%	10.08%	10.35%
13			Holland and Welsch		6.45%	9.27%	8.46%	21.20%
14		Regularisation	Lasso		9.38%	11.11%	9.80%	8.19%
15			Ridge Regression		9.38%	11.11%	9.80%	8.19%
16			Elastic Nets		9.38%	11.11%	9.80%	8.19%
17		Stepwise	Interaction	Bounded Steps	6.70%	4.75%	4.86%	6.03%
18				Unbounded Steps	6.69%	4.74%	4.84%	5.92%
19			Purequadratic	Bounded Steps	9.38%	11.11%	9.80%	8.19%
20				Unbounded Steps	6.68%	4.73%	4.82%	5.88%
21			Quadratic	Bounded Steps	6.70%	4.75%	4.86%	6.03%
22				Unbounded Steps	6.68%	4.73%	4.82%	5.88%
23			Polynomial	Bounded Steps	6.70%	4.75%	4.86%	6.03%
24				Unbounded Steps	82.81%	58.87%	49.97%	12.99%

Table H.15: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,7* (*cont.*)

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma}: Test Set T_o, o=5 *I-II-A-5*	Test Set T_o, o=6 *I-II-A-6*	Test Set T_u, u=2 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			9.03%	23.53%	5.92%	30.01%	**42.70%**	**0.77**
2		Bootstrap Aggregating			27.49%	46.92%	15.59%	91.72%	**65.30%**	**0.49**
3	Neural Network	Feedforward			5.73%	8.70%	1.96%	2.35%	**36.80%**	**1.09**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		41.00%	6.72%	4.34%	37.18%	**16.83%**	**0.87**
7			Andrews		41.00%	6.72%	4.34%	37.18%	**16.83%**	**0.87**
8			Cauchy M-estimators by Moore		40.87%	6.67%	4.28%	37.42%	**16.84%**	**0.87**
9			Fair by Rey		11.92%	7.65%	5.84%	29.65%	**11.61%**	**0.65**
10			Huber		17.99%	7.50%	5.32%	32.28%	**13.01%**	**0.67**
11			Logistic Regression		16.56%	7.46%	5.44%	31.20%	**12.50%**	**0.66**
12			Hinch and Talwar		9.64%	8.24%	6.23%	30.45%	**12.01%**	**0.63**
13			Holland and Welsch		40.92%	6.72%	4.33%	37.20%	**16.82%**	**0.87**
14		Regularisation	Lasso		5.97%	8.23%	6.22%	30.51%	**11.18%**	**0.72**
15			Ridge Regression		5.97%	8.23%	6.22%	30.51%	**11.18%**	**0.72**
16			Elastic Nets		5.97%	8.23%	6.22%	30.51%	**11.18%**	**0.72**
17		Stepwise	Interaction	Bounded Steps	4.23%	1.81%	2.44%	2.31%	**4.14%**	**0.43**
18				Unbounded Steps	4.05%	1.71%	2.23%	2.75%	**4.12%**	**0.43**
19			Purequadratic	Bounded Steps	5.97%	8.23%	6.22%	30.51%	**11.18%**	**0.72**
20				Unbounded Steps	3.96%	1.61%	2.11%	3.10%	**4.11%**	**0.43**
21			Quadratic	Bounded Steps	4.23%	1.81%	2.44%	2.31%	**4.14%**	**0.43**
22				Unbounded Steps	3.96%	1.61%	2.11%	3.10%	**4.11%**	**0.43**
23			Polynomial	Bounded Steps	4.23%	1.81%	2.44%	2.31%	**4.14%**	**0.43**
24				Unbounded Steps	3.91%	3.31%	2.02%	2.24%	**27.02%**	**1.19**

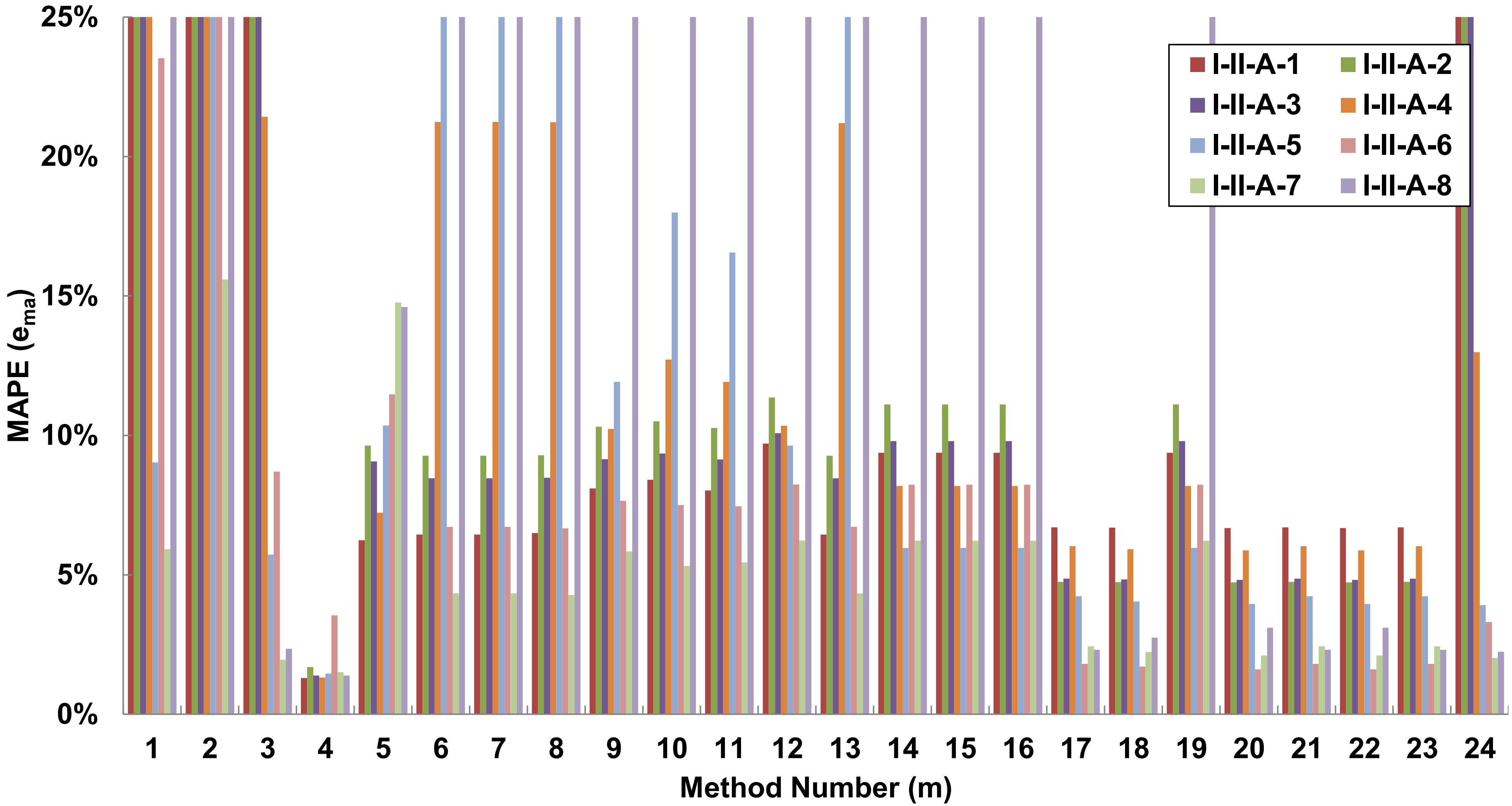


Figure H.15: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,7*

Table H.16: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,1*

	Method				$\mathbf{e_{ma}}$			
					Test Set $T_u, u=3$	Test Set $T_o, o12$	Test Set $T_u, u=2$	Test Set $T_o, o=2$
Number (m)	Class I	Class II	Class III	Class IV	*I-II-A-1*	*I-II-A-2*	*I-II-A-3*	*I-II-A-4*
1	Decision Tree	Boosting			4.10%	181.67%	2.77%	388.57%
2		Bootstrap Aggregating			11.58%	8.91%	4.74%	404.28%
3	Neural Network	Feedforward			2.93%	21.98%	2.87%	83.04%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		3.70%	5.93%	4.76%	14.76%
7			Andrews		3.70%	5.93%	4.76%	14.76%
8			Cauchy M-estimators by Moore		4.43%	7.42%	6.43%	7.13%
9			Fair by Rey		3.70%	5.61%	4.28%	19.81%
10			Huber		3.70%	5.77%	4.52%	18.25%
11			Logistic Regression		3.68%	5.73%	4.47%	18.10%
12			Hinch and Talwar		3.51%	5.25%	3.88%	19.27%
13			Holland and Welsch		3.71%	5.94%	4.77%	14.75%
14		Regularisation	Lasso		4.02%	4.51%	3.18%	30.35%
15			Ridge Regression		4.17%	5.83%	4.95%	33.58%
16			Elastic Nets		4.27%	4.46%	2.66%	33.97%
17		Stepwise	Interaction	Bounded Steps	3.62%	4.52%	4.53%	35.76%
18				Unbounded Steps	3.79%	4.00%	2.44%	36.08%
19			Purequadratic	Bounded Steps	4.10%	4.63%	2.71%	32.54%
20				Unbounded Steps	3.41%	3.80%	2.47%	22.59%
21			Quadratic	Bounded Steps	3.34%	3.97%	3.35%	25.31%
22				Unbounded Steps	3.42%	3.76%	2.36%	22.68%
23			Polynomial	Bounded Steps	2.96%	3.59%	2.07%	4.91%
24				Unbounded Steps	2.82%	3.45%	2.07%	9.27%

Table H.16: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,1* (*cont.*)

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma}: Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			825.09%	122.92%	158.51%	420.52%	**263.02%**	**1.04**
2		Bootstrap Aggregating			757.92%	87.68%	79.38%	6.69%	**170.15%**	**1.60**
3	Neural Network	Feedforward			95.50%	64.55%	54.34%	50.65%	**46.98%**	**0.74**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		31.91%	10.90%	36.33%	229.36%	**42.21%**	**1.82**
7			Andrews		31.91%	10.91%	36.34%	229.40%	**42.21%**	**1.82**
8			Cauchy M-estimators by Moore		16.10%	4.56%	15.97%	121.85%	**22.99%**	**1.75**
9			Fair by Rey		42.10%	18.29%	47.86%	286.72%	**53.55%**	**1.79**
10			Huber		41.09%	14.66%	41.79%	255.16%	**48.12%**	**1.77**
11			Logistic Regression		40.74%	14.59%	42.05%	256.51%	**48.23%**	**1.77**
12			Hinch and Talwar		41.14%	15.30%	48.29%	288.55%	**53.15%**	**1.82**
13			Holland and Welsch		31.90%	10.87%	36.23%	228.87%	**42.13%**	**1.81**
14		Regularisation	Lasso		63.65%	36.32%	78.40%	443.46%	**82.99%**	**1.79**
15			Ridge Regression		55.58%	34.33%	76.33%	434.15%	**81.12%**	**1.79**
16			Elastic Nets		62.13%	30.20%	71.89%	410.69%	**77.53%**	**1.77**
17		Stepwise	Interaction	Bounded Steps	80.92%	8.66%	8.51%	5.18%	**18.96%**	**1.43**
18				Unbounded Steps	70.14%	8.28%	7.81%	2.68%	**16.90%**	**1.43**
19			Purequadratic	Bounded Steps	57.31%	34.51%	76.25%	431.41%	**80.43%**	**1.79**
20				Unbounded Steps	30.96%	6.01%	6.27%	3.37%	**9.86%**	**1.09**
21			Quadratic	Bounded Steps	47.30%	7.10%	7.87%	8.74%	**13.37%**	**1.16**
22				Unbounded Steps	31.22%	5.95%	6.19%	2.53%	**9.76%**	**1.12**
23			Polynomial	Bounded Steps	5.00%	3.67%	3.93%	4.97%	**3.89%**	**0.27**
24				Unbounded Steps	25.88%	17.80%	19.54%	4.50%	**10.67%**	**0.86**

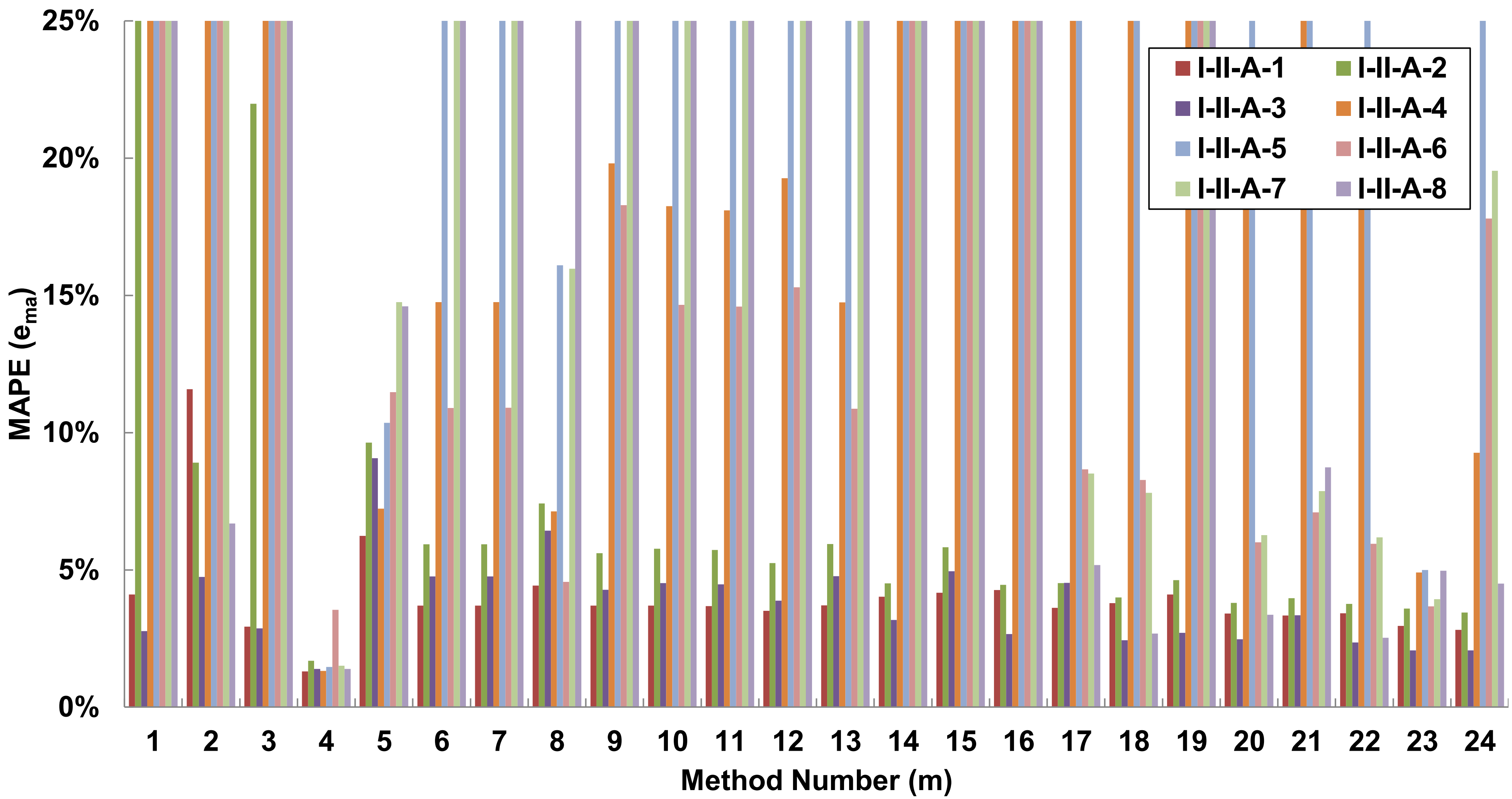


Figure H.16: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,1*

Table H.17: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,2*

Number (m)	Method				$\mathbf{e_{ma}}$			
					Test Set T_o, o=1	Test Set T_u, u=3	Test Set T_u, u=2	Test Set T_o, o=2
	Class I	Class II	Class III	Class IV	*I-II-A-1*	*I-II-A-2*	*I-II-A-3*	*I-II-A-4*
1	Decision Tree	Boosting			7.79%	5.13%	3.46%	342.23%
2		Bootstrap Aggregating						
3	Neural Network	Feedforward						
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.37%	4.05%	2.31%	45.36%
7			Andrews		5.38%	4.05%	2.31%	45.40%
8			Cauchy M-estimators by Moore		4.28%	3.77%	2.32%	39.99%
9			Fair by Rey		4.09%	3.80%	2.57%	38.94%
10			Huber		4.01%	3.77%	2.50%	37.66%
11			Logistic Regression		4.07%	3.77%	2.50%	37.72%
12			Hinch and Talwar		5.27%	4.06%	2.32%	44.75%
13			Holland and Welsch		5.31%	4.02%	2.29%	45.08%
14		Regularisation	Lasso					
15			Ridge Regression					
16			Elastic Nets					
17		Stepwise	Interaction	Bounded Steps	3.57%	3.32%	2.31%	21.51%
18				Unbounded Steps	6.04%	3.32%	2.11%	41.08%
19			Purequadratic	Bounded Steps	6.37%	3.95%	3.06%	61.63%
20				Unbounded Steps	5.99%	3.33%	2.04%	41.91%
21			Quadratic	Bounded Steps	3.63%	3.33%	2.24%	21.17%
22				Unbounded Steps	5.83%	3.33%	2.05%	40.16%
23			Polynomial	Bounded Steps	3.16%	3.28%	2.25%	10.53%
24				Unbounded Steps				

Table H.17: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,2* (*cont.*)

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			724.21%	85.72%	76.10%	23.30%	**158.49%**	**1.61**
2	Decision Tree	Bootstrap Aggregating							**#DIV/0!**	
3	Neural Network	Feedforward							**#DIV/0!**	
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		68.02%	29.25%	82.05%	458.78%	**86.90%**	**1.76**
7	Multiple Linear Regression	Robust	Andrews		68.01%	29.23%	82.03%	458.71%	**86.89%**	**1.76**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		74.53%	37.24%	93.80%	518.54%	**96.81%**	**1.79**
9	Multiple Linear Regression	Robust	Fair by Rey		80.00%	43.09%	101.11%	555.40%	**103.63%**	**1.80**
10	Multiple Linear Regression	Robust	Huber		78.00%	40.92%	98.95%	544.17%	**101.25%**	**1.80**
11	Multiple Linear Regression	Robust	Logistic Regression		78.15%	41.23%	99.18%	545.35%	**101.50%**	**1.80**
12	Multiple Linear Regression	Robust	Hinch and Talwar		68.13%	29.48%	82.75%	462.13%	**87.36%**	**1.77**
13	Multiple Linear Regression	Robust	Holland and Welsch		68.36%	29.53%	82.45%	461.05%	**87.26%**	**1.77**
14	Multiple Linear Regression	Regularisation	Lasso							
15	Multiple Linear Regression	Regularisation	Ridge Regression							
16	Multiple Linear Regression	Regularisation	Elastic Nets							
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	16.06%	4.16%	7.45%	36.15%	**11.82%**	**1.02**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	24.90%	10.71%	11.47%	35.91%	**16.94%**	**0.89**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	91.87%	57.60%	118.44%	640.99%	**122.99%**	**1.74**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	20.87%	4.30%	7.51%	37.10%	**15.38%**	**1.04**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	11.78%	4.29%	7.50%	37.07%	**11.38%**	**1.07**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	20.73%	6.08%	7.90%	38.84%	**15.62%**	**1.01**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	7.94%	4.37%	7.43%	37.11%	**9.51%**	**1.21**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps						

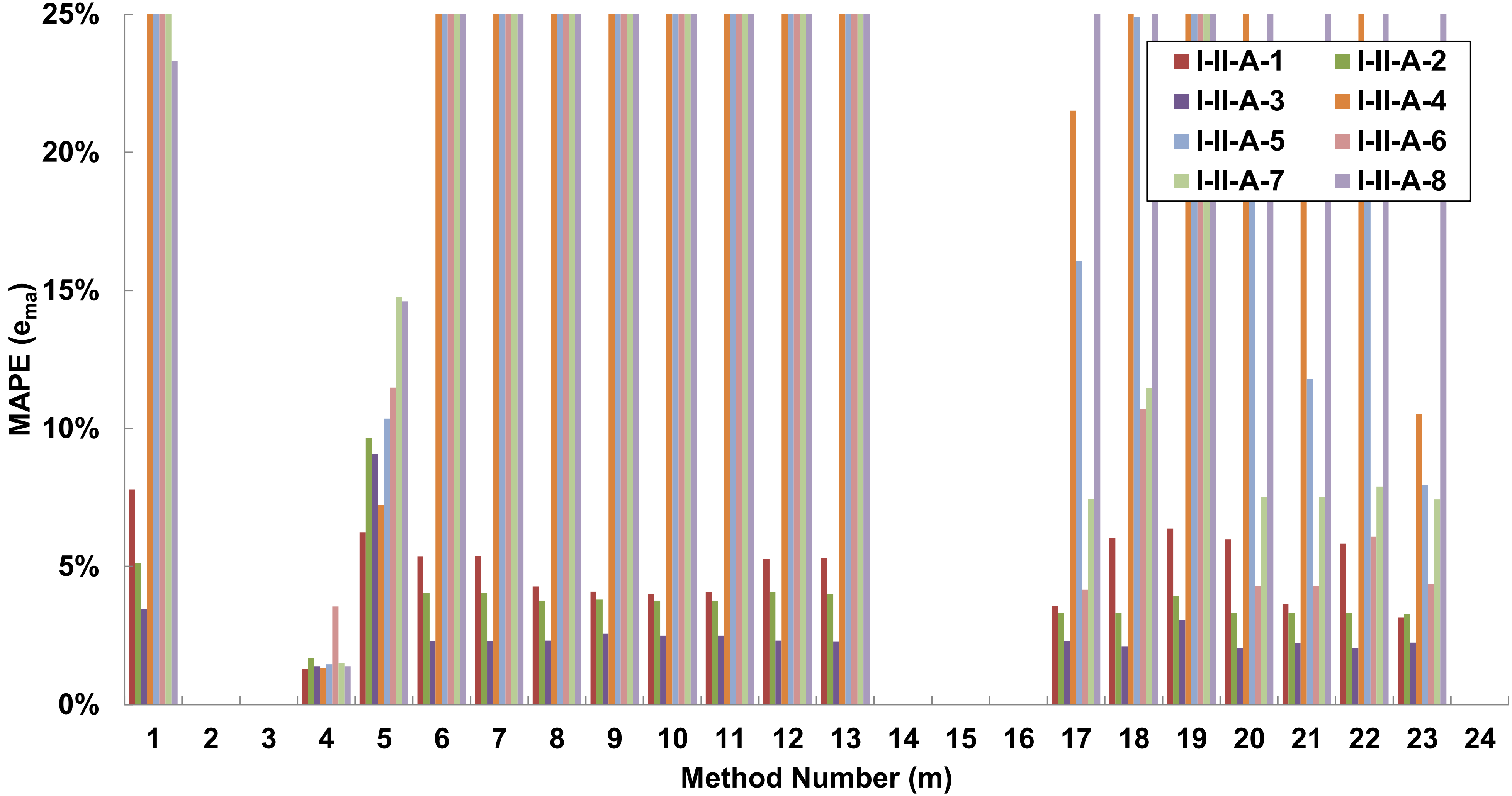


Figure H.17: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,2*

Table H.18: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4*

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ *I-II-A-1*	Test Set $T_o, o=1$ *I-II-A-2*	Test Set $T_u, u=2$ *I-II-A-3*	Test Set $T_u, u=3$ *I-II-A-4*
1	Decision Tree	Boosting			7.04%	37.50%	6.85%	4.17%
2	Decision Tree	Bootstrap Aggregating			20.39%	20.69%	6.24%	18.04%
3	Neural Network	Feedforward			5.19%	3.98%	1.90%	2.50%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.81%	9.26%	8.40%	6.05%
7	Multiple Linear Regression	Robust	Andrews		5.81%	9.26%	8.40%	6.05%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		5.81%	9.25%	8.39%	6.00%
9	Multiple Linear Regression	Robust	Fair by Rey		5.62%	8.82%	7.90%	4.45%
10	Multiple Linear Regression	Robust	Huber		5.58%	8.97%	8.05%	5.03%
11	Multiple Linear Regression	Robust	Logistic Regression		5.58%	8.96%	8.04%	4.98%
12	Multiple Linear Regression	Robust	Hinch and Talwar		5.81%	9.26%	8.39%	6.05%
13	Multiple Linear Regression	Robust	Holland and Welsch		5.81%	9.26%	8.40%	6.05%
14	Multiple Linear Regression	Regularisation	Lasso		5.87%	8.42%	7.37%	3.77%
15	Multiple Linear Regression	Regularisation	Ridge Regression		5.77%	8.62%	7.61%	5.00%
16	Multiple Linear Regression	Regularisation	Elastic Nets		5.87%	8.42%	7.37%	3.77%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	3.61%	5.38%	3.58%	5.40%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	3.91%	5.39%	3.51%	5.63%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	6.17%	8.88%	7.80%	4.39%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	3.83%	5.17%	3.37%	5.10%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	3.50%	5.47%	4.19%	3.49%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	3.86%	5.20%	3.37%	5.29%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	3.14%	3.56%	2.08%	2.69%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	67.14%	16.38%	1.99%	2.31%

Table H.18: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			39.10%	28.06%	109.85%	80.14%	**39.09%**	**0.97**
2		Bootstrap Aggregating			115.22%	75.96%	59.93%	7.36%	**40.48%**	**0.97**
3	Neural Network	Feedforward			3.72%	37.66%	32.81%	2.82%	**11.32%**	**1.31**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.90%	9.75%	11.57%	2.31%	**7.63%**	**0.38**
7			Andrews		7.90%	9.75%	11.57%	2.31%	**7.63%**	**0.38**
8			Cauchy M-estimators by Moore		7.79%	9.74%	11.53%	2.36%	**7.61%**	**0.37**
9			Fair by Rey		4.29%	7.24%	6.73%	32.49%	**9.69%**	**0.96**
10			Huber		5.86%	9.01%	9.65%	17.30%	**8.68%**	**0.45**
11			Logistic Regression		5.73%	8.77%	9.22%	17.89%	**8.65%**	**0.47**
12			Hinch and Talwar		7.89%	9.74%	11.55%	2.30%	**7.62%**	**0.38**
13			Holland and Welsch		7.89%	9.75%	11.57%	2.30%	**7.63%**	**0.38**
14		Regularisation	Lasso		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
15			Ridge Regression		10.05%	5.98%	8.91%	86.47%	**17.30%**	**1.62**
16			Elastic Nets		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
17		Stepwise	Interaction	Bounded Steps	17.03%	38.48%	37.56%	2.88%	**14.24%**	**1.08**
18				Unbounded Steps	20.33%	8.19%	8.70%	4.12%	**7.47%**	**0.74**
19			Purequadratic	Bounded Steps	6.32%	4.05%	7.98%	87.67%	**16.66%**	**1.73**
20				Unbounded Steps	17.68%	7.37%	8.00%	4.80%	**6.92%**	**0.67**
21			Quadratic	Bounded Steps	3.29%	2.69%	4.05%	44.95%	**8.95%**	**1.63**
22				Unbounded Steps	18.64%	7.73%	8.32%	4.62%	**7.13%**	**0.70**
23			Polynomial	Bounded Steps	3.23%	3.16%	3.45%	3.55%	**3.11%**	**0.16**
24				Unbounded Steps	7.32%	64.04%	80.68%	4.91%	**30.60%**	**1.10**

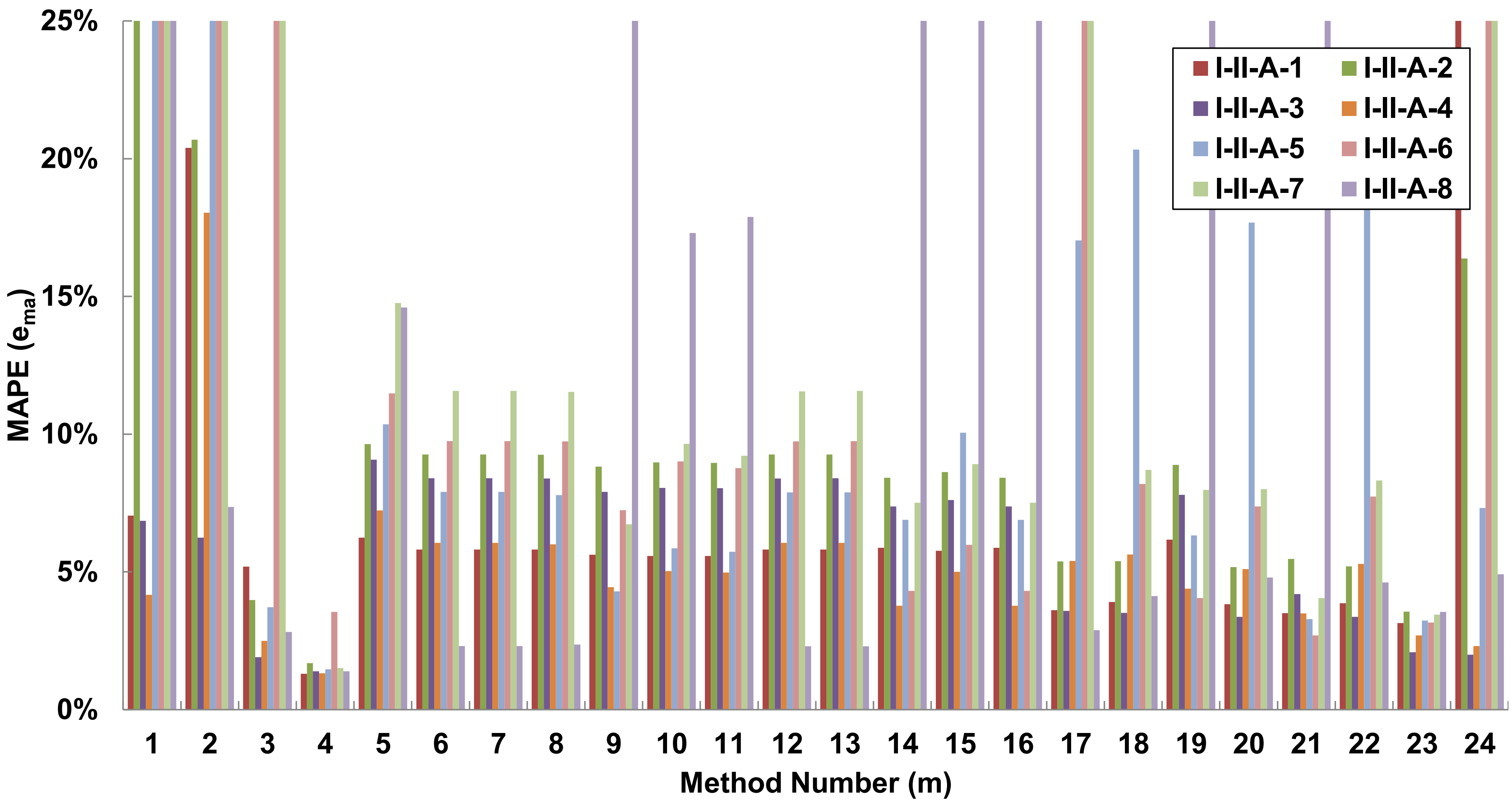


Figure H.18: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4*

Table H.19: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,5*

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ *I-II-A-1*	Test Set $T_o, o=2$ *I-II-A-2*	Test Set $T_u, u=2$ *I-II-A-3*	Test Set $T_o, o=3$ *I-II-A-4*
1	Decision Tree	Boosting			14.06%	34.35%	6.91%	7.03%
2	Decision Tree	Bootstrap Aggregating			35.14%	23.13%	17.74%	16.73%
3	Neural Network	Feedforward			242.19%	65.62%	1.84%	787.60%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.90%	8.93%	8.07%	8.75%
7	Multiple Linear Regression	Robust	Andrews		5.86%	8.99%	8.14%	8.19%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		5.76%	9.23%	8.35%	6.04%
9	Multiple Linear Regression	Robust	Fair by Rey		5.81%	8.85%	7.90%	6.14%
10	Multiple Linear Regression	Robust	Huber		5.71%	8.85%	7.91%	6.70%
11	Multiple Linear Regression	Robust	Logistic Regression		5.71%	8.85%	7.91%	6.56%
12	Multiple Linear Regression	Robust	Hinch and Talwar		6.31%	8.92%	7.93%	10.13%
13	Multiple Linear Regression	Robust	Holland and Welsch		5.90%	8.93%	8.07%	8.76%
14	Multiple Linear Regression	Regularisation	Lasso		7.65%	9.62%	8.36%	4.68%
15	Multiple Linear Regression	Regularisation	Ridge Regression		8.66%	9.79%	8.37%	9.64%
16	Multiple Linear Regression	Regularisation	Elastic Nets		7.65%	9.62%	8.36%	4.68%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	3.83%	3.48%	2.10%	3.80%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	3.83%	3.48%	2.10%	3.80%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	12.40%	10.09%	7.91%	7.59%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	3.80%	3.45%	2.05%	3.66%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	3.83%	3.47%	2.10%	3.79%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	3.80%	3.45%	2.05%	3.66%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	3.83%	3.47%	2.10%	3.79%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	140.63%	7.47%	2.02%	16.02%

Table H.19: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,5* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_u, u=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			4.38%	56.46%	96.48%	62.76%	**35.30%**	**0.96**
2		Bootstrap Aggregating			33.71%	83.31%	72.82%	4.57%	**35.89%**	**0.78**
3	Neural Network	Feedforward			2.95%	12.10%	15.24%	3.10%	**141.33%**	**1.94**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.13%	8.20%	7.48%	20.50%	**9.12%**	**0.53**
7			Andrews		5.64%	8.52%	8.31%	16.46%	**8.76%**	**0.38**
8			Cauchy M-estimators by Moore		7.78%	9.66%	11.44%	2.42%	**7.59%**	**0.37**
9			Fair by Rey		4.38%	7.59%	7.41%	29.79%	**9.73%**	**0.84**
10			Huber		4.70%	8.39%	8.16%	26.08%	**9.56%**	**0.71**
11			Logistic Regression		4.67%	8.25%	8.01%	26.17%	**9.52%**	**0.72**
12			Hinch and Talwar		4.57%	7.76%	6.27%	26.07%	**9.75%**	**0.70**
13			Holland and Welsch		5.15%	8.21%	7.52%	20.38%	**9.12%**	**0.52**
14		Regularisation	Lasso		4.95%	3.76%	5.14%	72.34%	**14.56%**	**1.61**
15			Ridge Regression		5.20%	3.77%	4.62%	71.22%	**15.16%**	**1.50**
16			Elastic Nets		4.95%	3.76%	5.14%	72.34%	**14.56%**	**1.61**
17		Stepwise	Interaction	Bounded Steps	3.21%	3.85%	4.40%	2.84%	**3.44%**	**0.21**
18				Unbounded Steps	3.21%	3.85%	4.40%	2.84%	**3.44%**	**0.21**
19			Purequadratic	Bounded Steps	4.63%	11.07%	10.63%	69.17%	**16.69%**	**1.28**
20				Unbounded Steps	2.99%	3.60%	4.20%	5.55%	**3.66%**	**0.27**
21			Quadratic	Bounded Steps	3.17%	3.82%	4.37%	2.81%	**3.42%**	**0.21**
22				Unbounded Steps	2.99%	3.60%	4.20%	5.55%	**3.66%**	**0.27**
23			Polynomial	Bounded Steps	3.17%	3.82%	4.37%	2.81%	**3.42%**	**0.21**
24				Unbounded Steps	2.91%	3.95%	4.94%	4.63%	**22.82%**	**2.09**

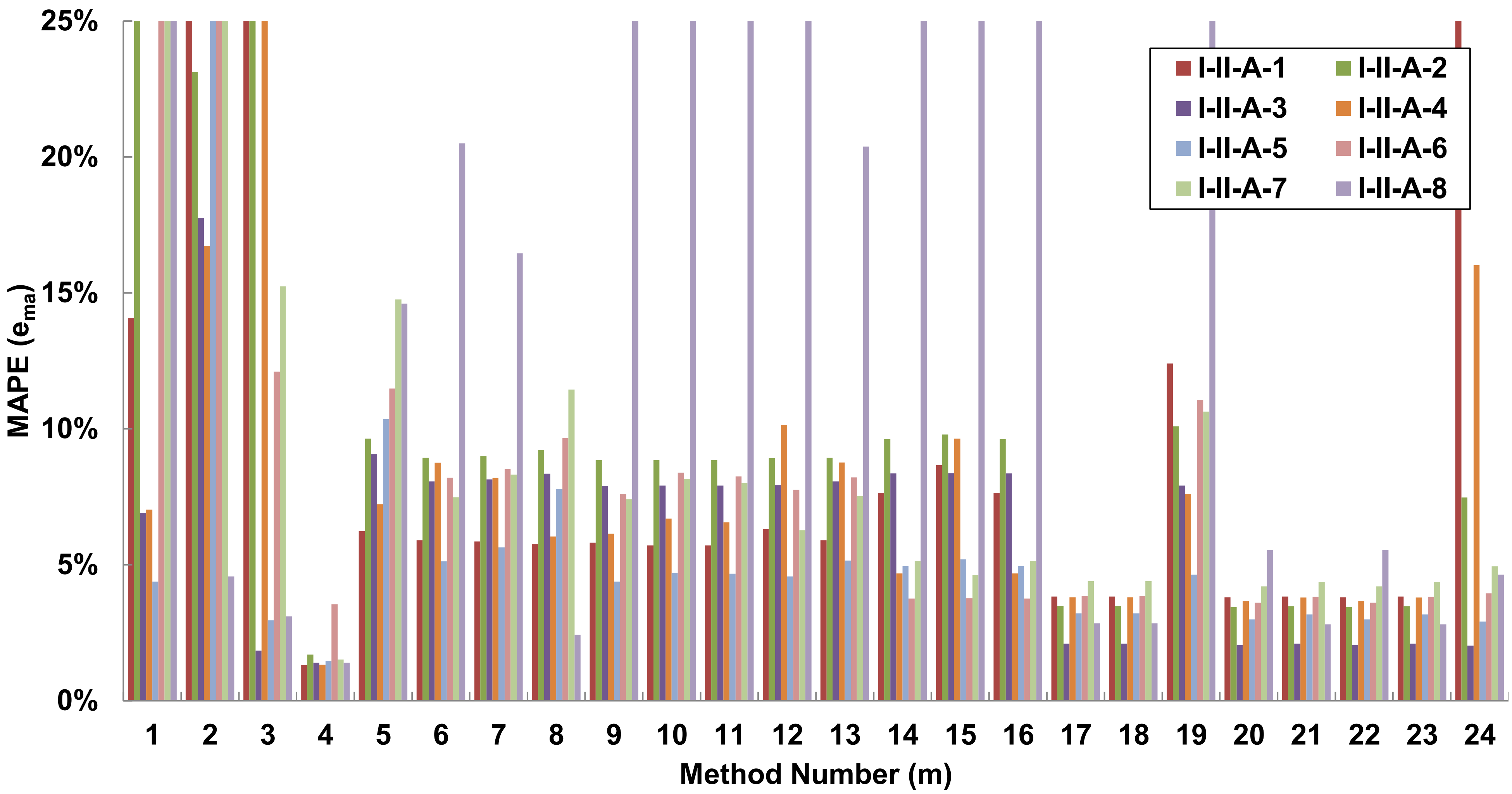


Figure H.19: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,5*

Table H.20: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,6*

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma}: Test Set $T_o, o=1$ *I-II-A-1*	Test Set $T_o, o=2$ *I-II-A-2*	Test Set $T_u, u=2$ *I-II-A-3*	Test Set $T_o, o=3$ *I-II-A-4*
1	Decision Tree	Boosting			13.90%	464.80%	6.65%	361.25%
2		Bootstrap Aggregating			22.07%	11.50%	5.70%	274.57%
3	Neural Network	Feedforward			75.60%	16.64%	1.75%	257.09%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.76%	9.24%	8.37%	6.14%
7			Andrews		5.76%	9.24%	8.37%	6.15%
8			Cauchy M-estimators by Moore		5.76%	9.24%	8.36%	6.12%
9			Fair by Rey		6.64%	9.22%	8.08%	16.47%
10			Huber		5.86%	9.18%	8.24%	8.30%
11			Logistic Regression		5.89%	9.18%	8.24%	8.70%
12			Hinch and Talwar		5.76%	9.24%	8.36%	6.14%
13			Holland and Welsch		5.77%	9.24%	8.37%	6.14%
14		Regularisation	Lasso		8.21%	9.38%	7.48%	47.46%
15			Ridge Regression		13.99%	10.24%	7.48%	32.95%
16			Elastic Nets		13.99%	10.24%	7.48%	32.95%
17		Stepwise	Interaction	Bounded Steps	6.52%	4.07%	2.02%	41.06%
18				Unbounded Steps	4.06%	3.82%	2.06%	27.46%
19			Purequadratic	Bounded Steps	13.32%	8.96%	7.41%	96.16%
20				Unbounded Steps	31.04%	7.57%	2.01%	156.05%
21			Quadratic	Bounded Steps	5.33%	4.04%	2.02%	38.72%
22				Unbounded Steps	7.98%	4.23%	2.01%	60.83%
23			Polynomial	Bounded Steps	17788.59%	359.52%	1.90%	3619.94%
24				Unbounded Steps	156.43%	19.27%	1.85%	692.02%

Table H.20: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,6* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=4 *I-II-A-5*	Test Set T_u, u=3 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			784.43%	3.64%	63.11%	77.95%	**221.97%**	**1.29**
2	Decision Tree	Bootstrap Aggregating			619.88%	30.23%	14.21%	54.58%	**129.09%**	**1.68**
3	Neural Network	Feedforward			409.28%	1.43%	3.83%	3.77%	**96.17%**	**1.60**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.87%	9.73%	11.54%	2.28%	**7.62%**	**0.38**
7	Multiple Linear Regression	Robust	Andrews		7.87%	9.73%	11.54%	2.28%	**7.62%**	**0.38**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		7.86%	9.71%	11.52%	2.28%	**7.61%**	**0.38**
9	Multiple Linear Regression	Robust	Fair by Rey		41.46%	9.26%	9.92%	8.47%	**13.69%**	**0.85**
10	Multiple Linear Regression	Robust	Huber		15.61%	9.33%	10.83%	3.28%	**8.83%**	**0.41**
11	Multiple Linear Regression	Robust	Logistic Regression		16.91%	9.33%	10.82%	3.30%	**9.05%**	**0.44**
12	Multiple Linear Regression	Robust	Hinch and Talwar		7.88%	9.74%	11.55%	2.29%	**7.62%**	**0.38**
13	Multiple Linear Regression	Robust	Holland and Welsch		7.87%	9.73%	11.54%	2.28%	**7.62%**	**0.38**
14	Multiple Linear Regression	Regularisation	Lasso		97.79%	6.51%	6.85%	91.60%	**34.41%**	**1.15**
15	Multiple Linear Regression	Regularisation	Ridge Regression		97.79%	6.51%	6.85%	91.60%	**33.43%**	**1.16**
16	Multiple Linear Regression	Regularisation	Elastic Nets		97.79%	6.51%	6.85%	91.60%	**33.43%**	**1.16**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	40.39%	1.76%	2.75%	2.50%	**12.63%**	**1.38**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	46.52%	1.83%	2.49%	2.34%	**11.32%**	**1.47**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	86.25%	4.05%	6.04%	89.05%	**38.91%**	**1.10**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	113.85%	1.93%	2.25%	7.76%	**40.31%**	**1.49**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	38.81%	1.76%	2.76%	2.48%	**11.99%**	**1.38**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	113.85%	1.93%	2.25%	7.76%	**25.11%**	**1.63**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	2076.79%	1.75%	2.78%	2.56%	**2981.73%**	**2.06**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	1338.78%	1.37%	3.37%	4.65%	**277.22%**	**1.77**

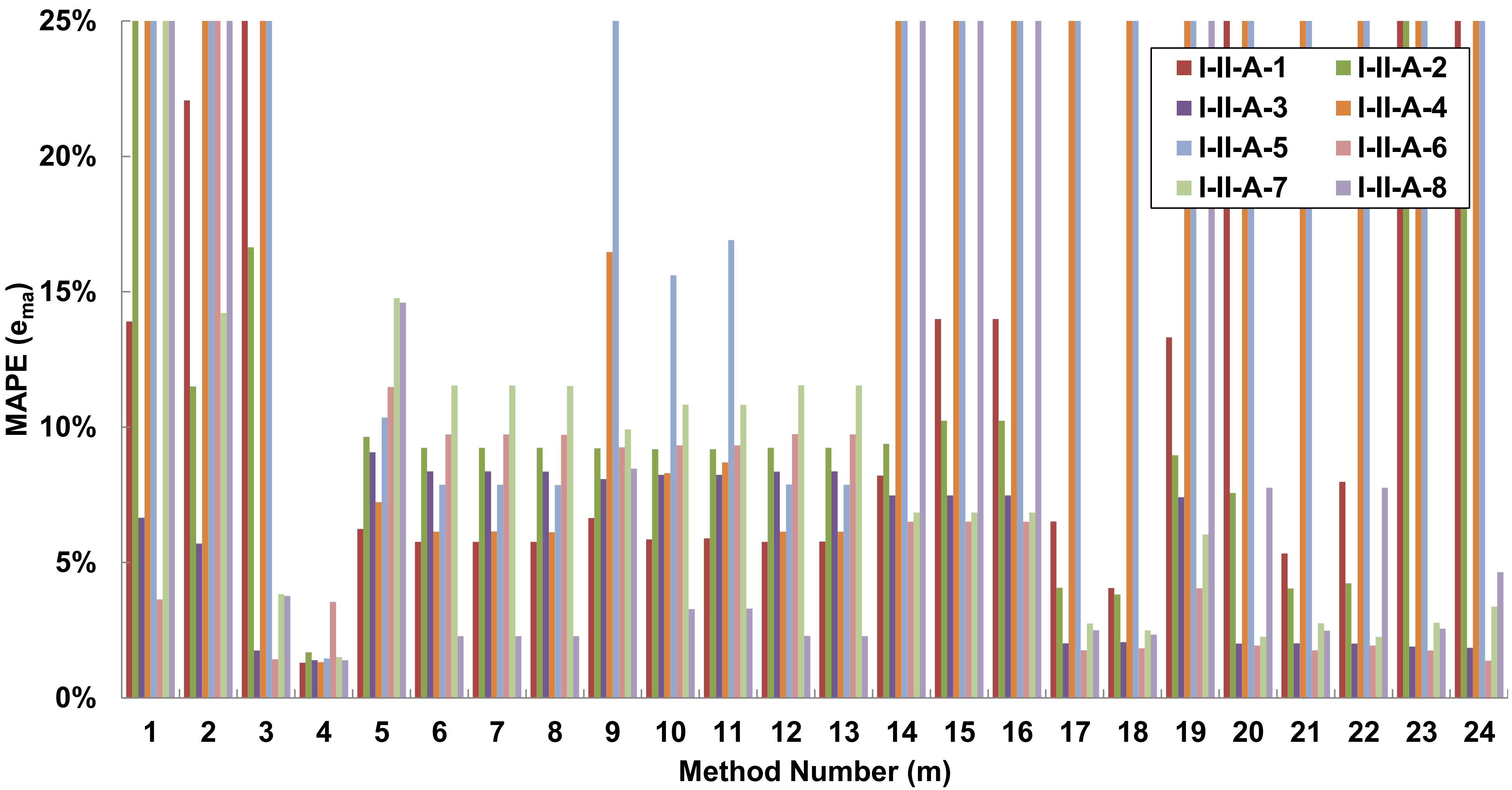


Figure H.20: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,6*

Table H.21: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,7*

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ *I-II-A-1*	Test Set $T_o, o=2$ *I-II-A-2*	Test Set $T_u, u=2$ *I-II-A-3*	Test Set $T_o, o=3$ *I-II-A-4*
1	Decision Tree	Boosting			14.02%	49.72%	6.87%	404.69%
2	Decision Tree	Bootstrap Aggregating			22.58%	11.80%	6.07%	291.79%
3	Neural Network	Feedforward			20.68%	18.03%	1.84%	302.19%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		6.83%	9.97%	9.30%	31.09%
7	Multiple Linear Regression	Robust	Andrews		6.86%	10.04%	9.39%	32.01%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		6.01%	8.67%	7.77%	10.43%
9	Multiple Linear Regression	Robust	Fair by Rey		8.18%	9.08%	7.56%	23.96%
10	Multiple Linear Regression	Robust	Huber		7.90%	8.98%	7.56%	23.06%
11	Multiple Linear Regression	Robust	Logistic Regression		7.70%	8.96%	7.57%	21.84%
12	Multiple Linear Regression	Robust	Hinch and Talwar		9.33%	10.63%	9.13%	8.25%
13	Multiple Linear Regression	Robust	Holland and Welsch		5.99%	8.68%	7.79%	10.49%
14	Multiple Linear Regression	Regularisation	Lasso		13.71%	10.44%	7.74%	34.03%
15	Multiple Linear Regression	Regularisation	Ridge Regression		13.71%	10.44%	7.74%	34.03%
16	Multiple Linear Regression	Regularisation	Elastic Nets		13.71%	10.44%	7.74%	34.03%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	4.04%	3.80%	2.06%	26.70%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	4.04%	3.80%	2.06%	26.70%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	9.80%	9.81%	7.68%	45.42%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	6.89%	3.96%	2.01%	50.41%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	5.90%	3.93%	2.04%	47.34%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	8.22%	4.27%	2.01%	61.21%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	16915.68%	298.36%	1.85%	2813.65%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	144.03%	16.87%	1.86%	691.51%

Table H.21: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,7* (*cont.*)

Number (m)	Method				e_{ma}					
					Test Set T_o, o=4	Test Set T_o, o=5	Test Set T_u, u=3	Test Set T_u, u=1		
	Class I	Class II	Class III	Class IV	*I-II-A-5*	*I-II-A-6*	*I-II-A-7*	*I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			873.93%	30.62%	3.73%	65.27%	**181.11%**	**1.71**
2		Bootstrap Aggregating			654.30%	43.19%	10.23%	60.82%	**137.60%**	**1.67**
3	Neural Network	Feedforward			66.39%	3.49%	1.97%	2.80%	**52.17%**	**1.98**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		70.06%	6.71%	4.33%	37.13%	**21.93%**	**1.05**
7			Andrews		72.78%	6.71%	4.33%	37.12%	**22.41%**	**1.06**
8			Cauchy M-estimators by Moore		4.36%	6.67%	4.27%	37.50%	**10.71%**	**1.03**
9			Fair by Rey		67.16%	7.27%	4.54%	38.90%	**20.83%**	**1.06**
10			Huber		64.41%	7.32%	4.68%	37.23%	**20.14%**	**1.04**
11			Logistic Regression		60.54%	7.24%	4.69%	36.78%	**19.42%**	**1.02**
12			Hinch and Talwar		15.53%	8.26%	6.24%	30.45%	**12.23%**	**0.64**
13			Holland and Welsch		3.79%	6.74%	4.35%	37.32%	**10.64%**	**1.03**
14		Regularisation	Lasso		102.37%	6.81%	4.64%	76.88%	**32.08%**	**1.16**
15			Ridge Regression		102.37%	6.81%	4.64%	76.88%	**32.08%**	**1.16**
16			Elastic Nets		102.37%	6.81%	4.64%	76.88%	**32.08%**	**1.16**
17		Stepwise	Interaction	Bounded Steps	45.32%	1.83%	2.46%	2.61%	**11.10%**	**1.46**
18				Unbounded Steps	45.32%	1.83%	2.46%	2.61%	**11.10%**	**1.46**
19			Purequadratic	Bounded Steps	91.23%	5.38%	3.75%	73.10%	**30.77%**	**1.13**
20				Unbounded Steps	114.86%	1.98%	2.17%	6.29%	**23.57%**	**1.71**
21			Quadratic	Bounded Steps	86.62%	1.68%	2.37%	4.77%	**19.33%**	**1.62**
22				Unbounded Steps	114.86%	1.98%	2.17%	6.29%	**25.13%**	**1.65**
23			Polynomial	Bounded Steps	1969.30%	2.63%	2.38%	4.21%	**2751.01%**	**2.12**
24				Unbounded Steps	1342.01%	2.34%	1.98%	3.99%	**275.57%**	**1.78**

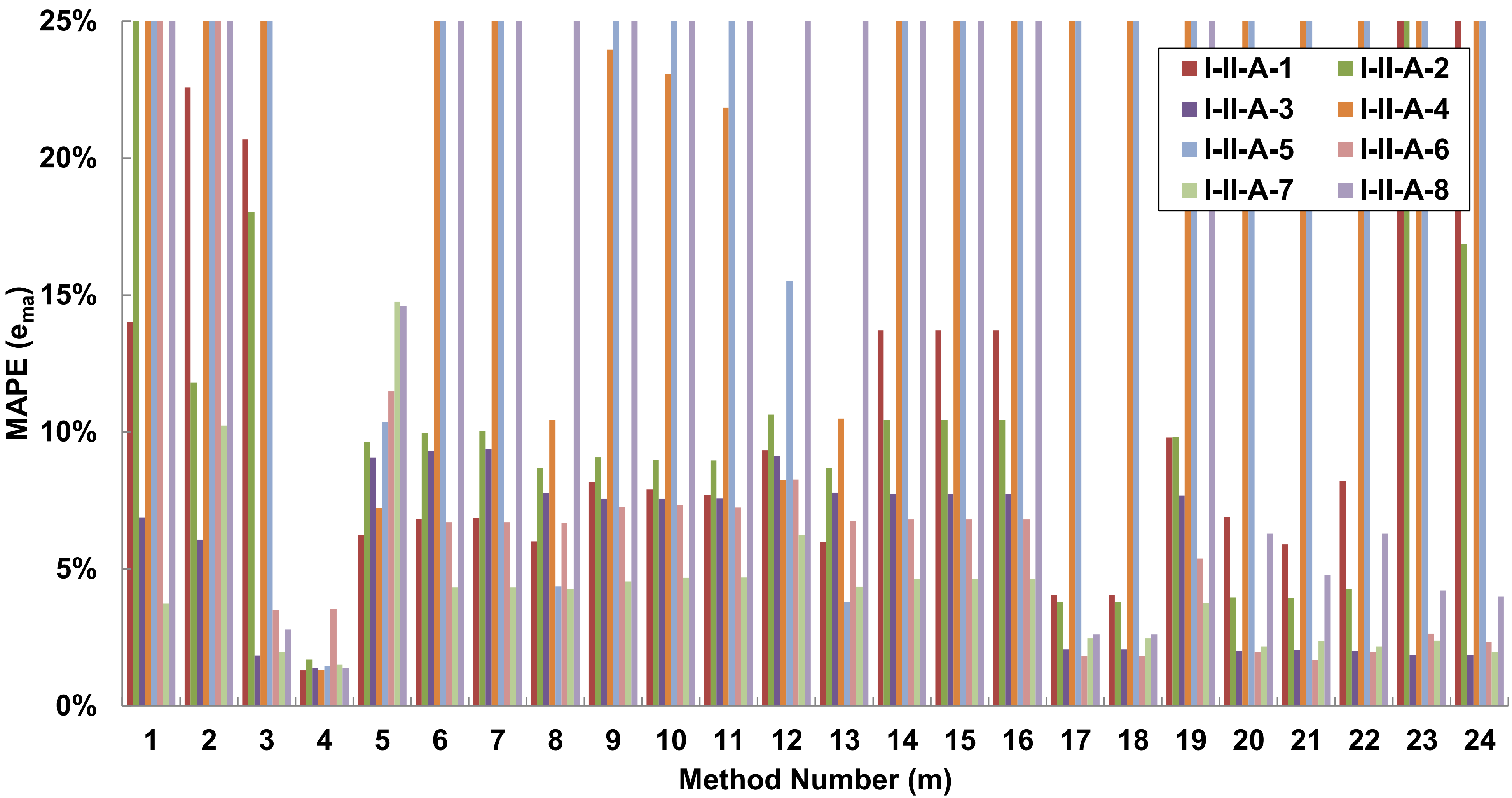


Figure H.21: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,7*

Table H.22: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,1*

Number (m)	Method Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_u, u=4$ *I-II-A-1*	Test Set $T_o, o=1$ *I-II-A-2*	Test Set $T_u, u=2$ *I-II-A-3*	Test Set $T_u, u=3$ *I-II-A-4*
1	Decision Tree	Boosting			4.15%	42.78%	3.35%	24.18%
2		Bootstrap Aggregating			12.63%	27.39%	7.51%	95.10%
3	Neural Network	Feedforward			3.87%	4.26%	2.41%	6.66%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.12%	8.38%	7.63%	3.64%
7			Andrews		5.12%	8.38%	7.63%	3.64%
8			Cauchy M-estimators by Moore		5.66%	9.10%	8.32%	5.63%
9			Fair by Rey		4.12%	6.57%	5.36%	10.16%
10			Huber		4.27%	7.03%	5.95%	7.37%
11			Logistic Regression		4.25%	7.00%	5.91%	7.46%
12			Hinch and Talwar		5.79%	9.27%	8.58%	6.07%
13			Holland and Welsch		5.13%	8.40%	7.63%	3.65%
14		Regularisation	Lasso		4.14%	5.01%	3.51%	23.32%
15			Ridge Regression		4.10%	5.66%	4.14%	24.10%
16			Elastic Nets		4.14%	5.01%	3.51%	23.32%
17		Stepwise	Interaction	Bounded Steps	3.52%	4.79%	2.75%	11.79%
18				Unbounded Steps	3.55%	4.71%	2.69%	12.10%
19			Purequadratic	Bounded Steps	4.16%	5.20%	3.49%	23.27%
20				Unbounded Steps	3.36%	4.01%	2.69%	9.91%
21			Quadratic	Bounded Steps	3.24%	3.60%	2.17%	11.88%
22				Unbounded Steps	3.36%	4.01%	2.69%	9.91%
23			Polynomial	Bounded Steps	2.94%	3.63%	2.13%	2.78%
24				Unbounded Steps	2.82%	3.47%	2.06%	2.55%

Table H.22: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,1* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=2 *I-II-A-5*	Test Set T_o, o=3 *I-II-A-6*	Test Set T_o, o=4 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			57.51%	23.30%	67.74%	345.44%	**71.06%**	**1.59**
2		Bootstrap Aggregating			123.08%	76.77%	61.27%	7.42%	**51.40%**	**0.86**
3	Neural Network	Feedforward			5.39%	91.79%	123.47%	390.80%	**78.58%**	**1.71**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.79%	5.75%	6.55%	60.14%	**12.88%**	**1.49**
7			Andrews		5.80%	5.75%	6.56%	60.20%	**12.89%**	**1.49**
8			Cauchy M-estimators by Moore		6.80%	9.89%	11.26%	10.82%	**8.44%**	**0.26**
9			Fair by Rey		21.92%	9.93%	28.26%	189.82%	**34.52%**	**1.83**
10			Huber		15.93%	7.21%	21.28%	152.08%	**27.64%**	**1.83**
11			Logistic Regression		16.29%	6.88%	21.22%	152.72%	**27.72%**	**1.83**
12			Hinch and Talwar		7.82%	10.02%	11.75%	2.47%	**7.72%**	**0.38**
13			Holland and Welsch		5.70%	5.85%	6.42%	59.01%	**12.72%**	**1.47**
14		Regularisation	Lasso		49.83%	27.49%	61.23%	358.07%	**66.58%**	**1.80**
15			Ridge Regression		44.69%	26.33%	59.94%	352.15%	**65.14%**	**1.81**
16			Elastic Nets		49.83%	27.49%	61.23%	358.07%	**66.58%**	**1.80**
17		Stepwise	Interaction	Bounded Steps	7.13%	20.76%	20.47%	2.63%	**9.23%**	**0.83**
18				Unbounded Steps	7.21%	21.16%	20.59%	14.20%	**10.78%**	**0.69**
19			Purequadratic	Bounded Steps	45.02%	26.88%	60.23%	351.36%	**64.95%**	**1.81**
20				Unbounded Steps	4.81%	43.55%	41.65%	12.90%	**15.36%**	**1.12**
21			Quadratic	Bounded Steps	20.74%	13.82%	14.59%	13.99%	**10.50%**	**0.64**
22				Unbounded Steps	4.81%	43.55%	41.65%	12.90%	**15.36%**	**1.12**
23			Polynomial	Bounded Steps	3.34%	3.07%	3.19%	7.13%	**3.53%**	**0.43**
24				Unbounded Steps	3.60%	11.95%	13.41%	3.02%	**5.36%**	**0.85**

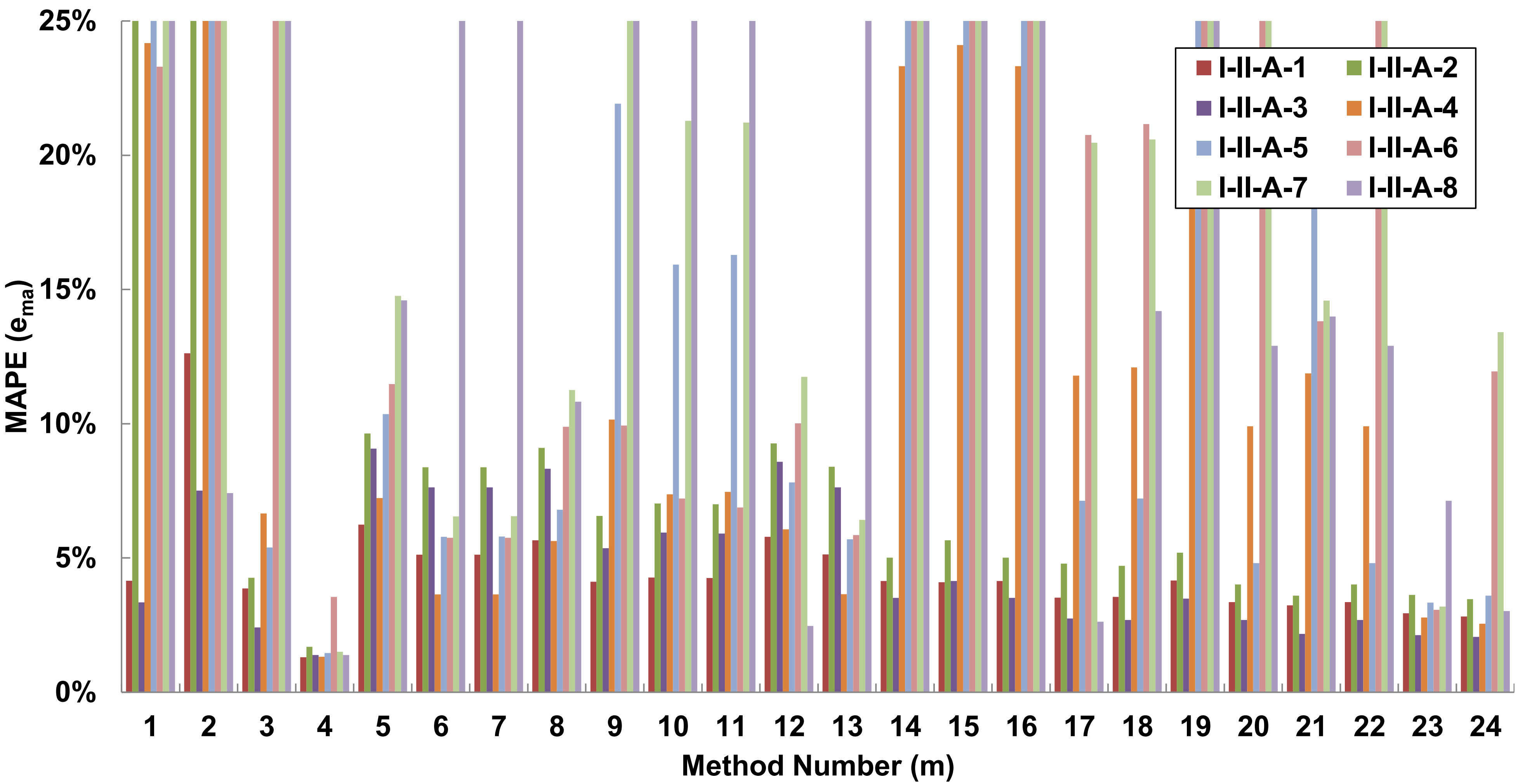


Figure H.22: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,1*

Table H.23: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,2*

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1: *I-II-A-1*	Test Set T_u, u=4: *I-II-A-2*	Test Set T_u, u=2: *I-II-A-3*	Test Set T_u, u=3: *I-II-A-4*
1	Decision Tree	Boosting			26.65%	4.98%	8.45%	102.62%
2	Decision Tree	Bootstrap Aggregating						
3	Neural Network	Feedforward						
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		3.55%	4.05%	2.58%	31.44%
7			Andrews		3.55%	4.05%	2.59%	31.43%
8			Cauchy M-estimators by Moore		3.76%	3.77%	2.36%	35.68%
9			Fair by Rey		4.07%	3.80%	2.56%	38.86%
10			Huber		3.97%	3.77%	2.50%	37.65%
11			Logistic Regression		3.98%	3.77%	2.51%	37.77%
12			Hinch and Talwar		3.57%	4.06%	2.59%	31.69%
13			Holland and Welsch		3.54%	4.02%	2.54%	31.56%
14		Regularisation	Lasso					
15			Ridge Regression					
16			Elastic Nets					
17		Stepwise	Interaction	Bounded Steps	3.92%	3.32%	2.43%	13.00%
18				Unbounded Steps	3.91%	3.32%	2.44%	13.12%
19			Purequadratic	Bounded Steps	4.98%	3.95%	2.93%	46.90%
20				Unbounded Steps	5.38%	3.32%	2.45%	14.24%
21			Quadratic	Bounded Steps	5.42%	3.32%	2.58%	14.27%
22				Unbounded Steps	5.36%	3.32%	2.75%	12.62%
23			Polynomial	Bounded Steps	2.92%	3.28%	2.18%	5.37%
24				Unbounded Steps				

Table H.23: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,2* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=2 *I-II-A-5*	Test Set T_o, o=3 *I-II-A-6*	Test Set T_o, o=4 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			175.96%	75.60%	59.59%	19.65%	**59.19%**	**0.99**
2		Bootstrap Aggregating								
3	Neural Network	Feedforward								
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		63.49%	29.20%	81.97%	458.42%	**84.34%**	**1.83**
7			Andrews		63.46%	29.19%	81.96%	458.33%	**84.32%**	**1.83**
8			Cauchy M-estimators by Moore		73.72%	37.03%	93.48%	516.91%	**95.84%**	**1.81**
9			Fair by Rey		79.72%	42.92%	100.76%	553.66%	**103.29%**	**1.80**
10			Huber		77.75%	40.75%	98.62%	542.49%	**100.94%**	**1.80**
11			Logistic Regression		77.93%	41.06%	98.84%	543.68%	**101.19%**	**1.80**
12			Hinch and Talwar		63.90%	29.42%	82.64%	461.61%	**84.94%**	**1.83**
13			Holland and Welsch		63.90%	29.44%	82.31%	460.31%	**84.70%**	**1.82**
14		Regularisation	Lasso							
15			Ridge Regression							
16			Elastic Nets							
17		Stepwise	Interaction	Bounded Steps	10.41%	206.25%	194.30%	35.59%	**58.65%**	**1.50**
18				Unbounded Steps	10.23%	206.21%	194.10%	35.71%	**58.63%**	**1.50**
19			Purequadratic	Bounded Steps	90.87%	57.23%	117.84%	638.29%	**120.37%**	**1.77**
20				Unbounded Steps	12.66%	12.79%	12.43%	54.69%	**14.75%**	**1.14**
21			Quadratic	Bounded Steps	12.66%	42.16%	38.54%	54.54%	**21.69%**	**0.94**
22				Unbounded Steps	9.76%	13.23%	13.37%	38.01%	**12.30%**	**0.92**
23			Polynomial	Bounded Steps	8.08%	4.76%	7.57%	43.16%	**9.67%**	**1.42**
24				Unbounded Steps						

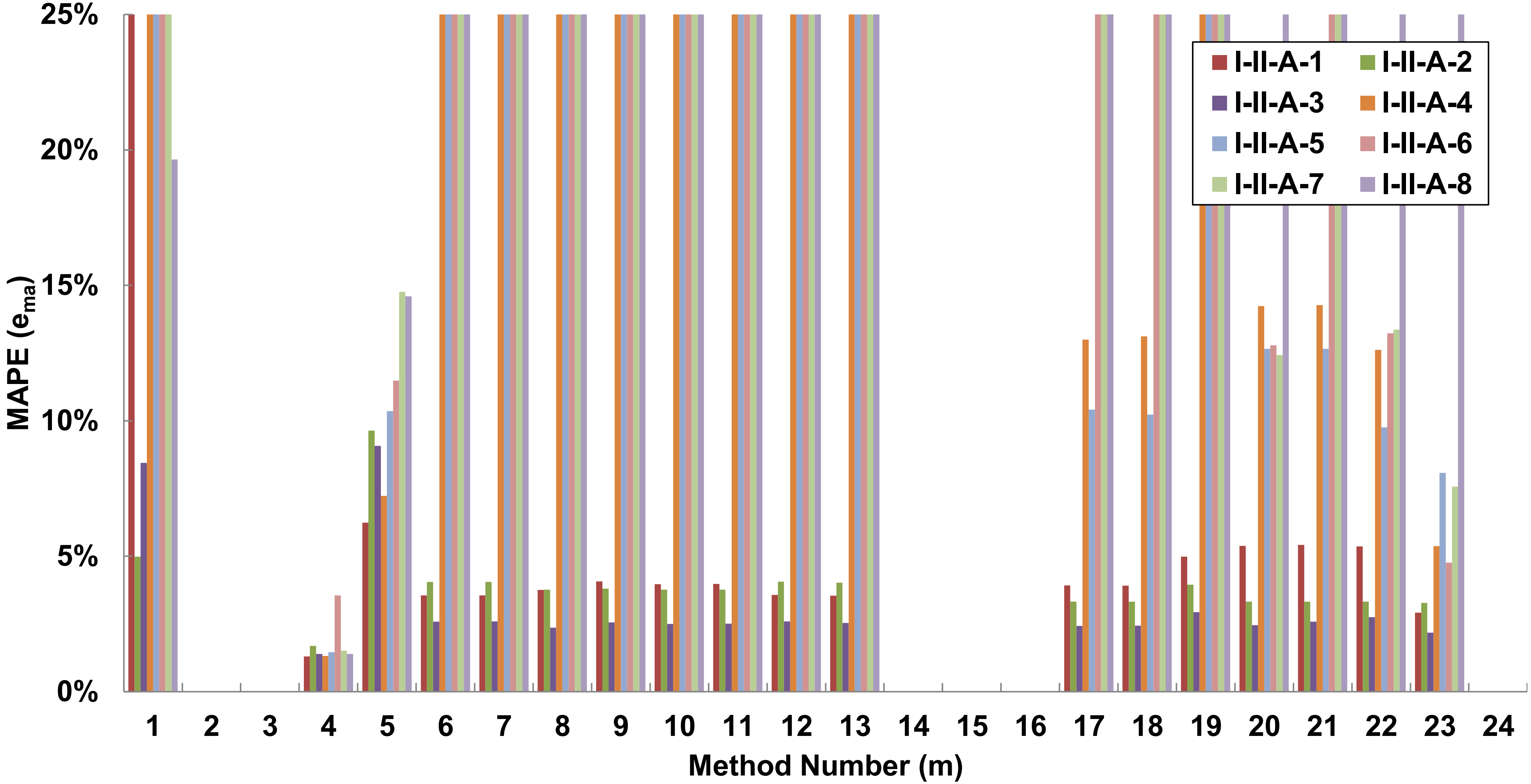


Figure H.23: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,2*

Table H.24: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set I-II-A-8,3,4,5

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1 *I-II-A-1*	Test Set T_o, o=2 *I-II-A-2*	Test Set T_u, u=4 *I-II-A-3*	Test Set T_u, u=3 *I-II-A-4*
1	Decision Tree	Boosting			7.22%	29.73%	7.00%	4.09%
2		Bootstrap Aggregating			25.15%	14.74%	5.39%	14.42%
3	Neural Network	Feedforward			4.20%	4.09%	1.85%	2.52%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.39%	8.71%	7.83%	3.95%
7			Andrews		5.39%	8.71%	7.82%	3.95%
8			Cauchy M-estimators by Moore		5.54%	8.89%	8.02%	4.49%
9			Fair by Rey		5.55%	8.70%	7.75%	4.07%
10			Huber		5.39%	8.67%	7.73%	4.10%
11			Logistic Regression		5.40%	8.67%	7.72%	4.08%
12			Hinch and Talwar		5.38%	8.70%	7.81%	3.93%
13			Holland and Welsch		5.56%	8.93%	8.06%	4.63%
14		Regularisation	Lasso		6.01%	8.61%	7.57%	3.77%
15			Ridge Regression		5.90%	8.73%	7.71%	4.09%
16			Elastic Nets		6.01%	8.61%	7.57%	3.77%
17		Stepwise	Interaction	Bounded Steps	3.83%	5.88%	4.11%	3.57%
18				Unbounded Steps	3.39%	5.50%	4.23%	3.57%
19			Purequadratic	Bounded Steps	6.37%	9.07%	7.98%	4.15%
20				Unbounded Steps	3.40%	5.22%	3.93%	3.40%
21			Quadratic	Bounded Steps	3.50%	5.46%	4.18%	3.52%
22				Unbounded Steps	3.40%	5.22%	3.93%	3.40%
23			Polynomial	Bounded Steps	3.17%	3.54%	2.07%	2.68%
24				Unbounded Steps	46.66%	15.52%	2.05%	2.36%

Table H.24: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,5* (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_u, u=4 *I-II-A-5*	Test Set T_o, o=3 *I-II-A-6*	Test Set T_o, o=4 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			4.90%	53.69%	88.66%	70.10%	**33.17%**	**1.01**
2	Decision Tree	Bootstrap Aggregating			14.02%	71.16%	72.75%	4.84%	**27.81%**	**1.01**
3	Neural Network	Feedforward			3.21%	70.98%	62.65%	4.85%	**19.29%**	**1.53**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		3.86%	6.88%	4.55%	36.44%	**9.70%**	**1.13**
7	Multiple Linear Regression	Robust	Andrews		3.85%	6.86%	4.52%	36.66%	**9.72%**	**1.14**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		4.61%	8.00%	6.83%	24.63%	**8.88%**	**0.74**
9	Multiple Linear Regression	Robust	Fair by Rey		3.50%	6.69%	5.58%	41.62%	**10.43%**	**1.22**
10	Multiple Linear Regression	Robust	Huber		3.77%	7.68%	6.45%	37.75%	**10.19%**	**1.11**
11	Multiple Linear Regression	Robust	Logistic Regression		3.70%	7.43%	6.09%	37.89%	**10.12%**	**1.12**
12	Multiple Linear Regression	Robust	Hinch and Talwar		3.83%	6.81%	4.45%	37.08%	**9.75%**	**1.15**
13	Multiple Linear Regression	Robust	Holland and Welsch		4.92%	8.11%	7.17%	22.36%	**8.72%**	**0.66**
14	Multiple Linear Regression	Regularisation	Lasso		5.37%	4.28%	5.51%	72.02%	**14.14%**	**1.66**
15	Multiple Linear Regression	Regularisation	Ridge Regression		5.05%	5.77%	7.09%	74.91%	**14.91%**	**1.63**
16	Multiple Linear Regression	Regularisation	Elastic Nets		5.37%	4.28%	5.51%	72.02%	**14.14%**	**1.66**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	6.53%	76.65%	73.26%	3.61%	**22.18%**	**1.47**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	6.14%	7.47%	8.00%	3.76%	**5.26%**	**0.34**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	4.97%	3.95%	6.02%	75.33%	**14.73%**	**1.67**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	5.80%	7.11%	7.66%	5.68%	**5.28%**	**0.31**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	3.29%	2.74%	3.98%	44.19%	**8.86%**	**1.61**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	5.80%	7.11%	7.66%	5.68%	**5.28%**	**0.31**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	3.22%	3.16%	3.41%	3.89%	**3.14%**	**0.18**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	2.92%	26.66%	18.27%	3.77%	**14.78%**	**1.07**

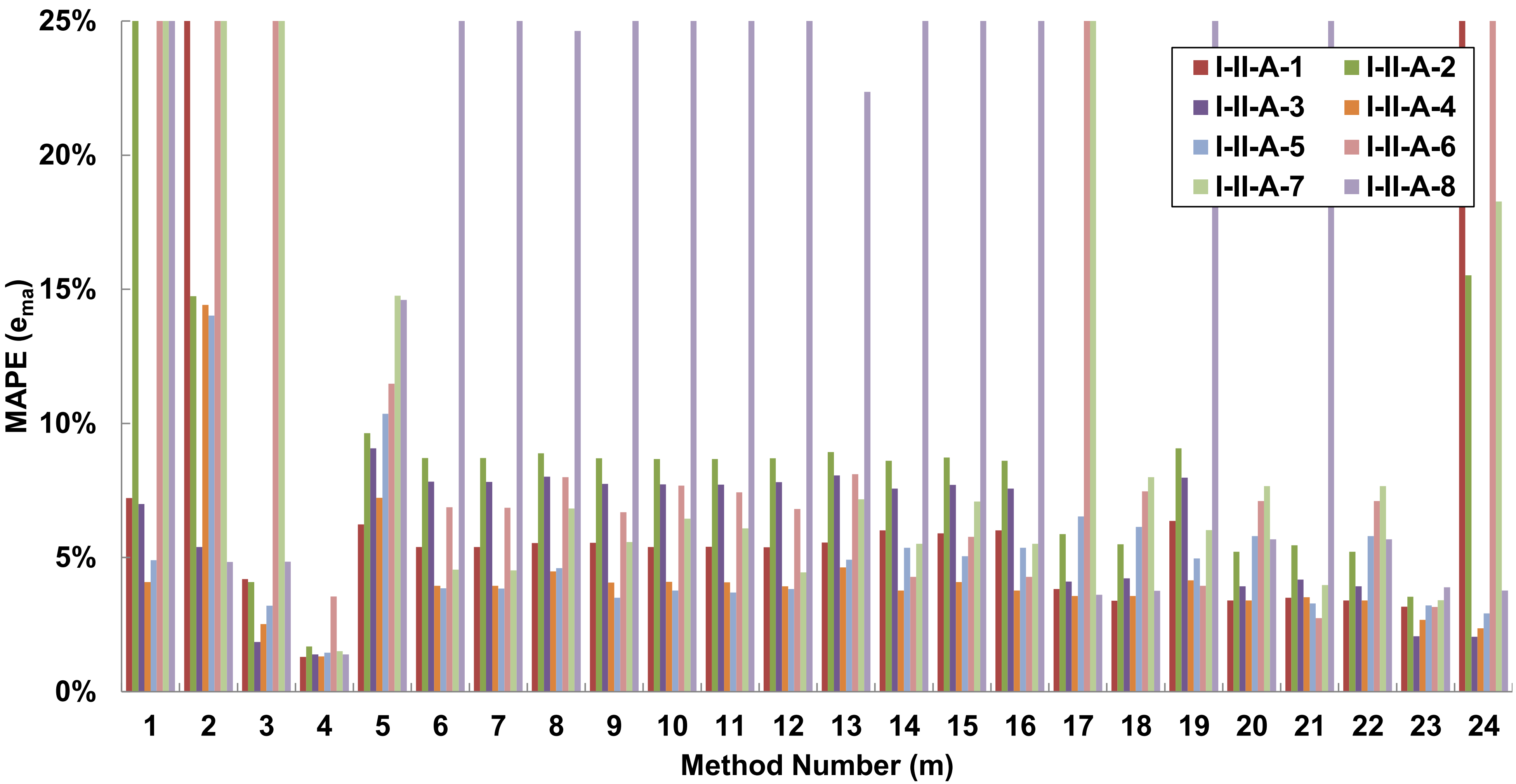


Figure H.24: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,5*

Table H.25: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,6*

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1 *I-II-A-1*	Test Set T_o, o=2 *I-II-A-2*	Test Set T_u, u=2 *I-II-A-3*	Test Set T_u, u=3 *I-II-A-4*
1	Decision Tree	Boosting			7.02%	28.71%	6.82%	4.18%
2	Decision Tree	Bootstrap Aggregating			23.74%	23.30%	9.02%	23.45%
3	Neural Network	Feedforward			4.08%	3.73%	2.02%	2.93%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.81%	9.26%	8.40%	6.06%
7	Multiple Linear Regression	Robust	Andrews		5.81%	9.26%	8.40%	6.06%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		5.81%	9.25%	8.39%	6.00%
9	Multiple Linear Regression	Robust	Fair by Rey		5.61%	8.76%	7.82%	4.21%
10	Multiple Linear Regression	Robust	Huber		5.54%	8.89%	7.97%	4.73%
11	Multiple Linear Regression	Robust	Logistic Regression		5.54%	8.87%	7.94%	4.66%
12	Multiple Linear Regression	Robust	Hinch and Talwar		5.30%	8.58%	7.68%	3.64%
13	Multiple Linear Regression	Robust	Holland and Welsch		5.81%	9.26%	8.40%	6.05%
14	Multiple Linear Regression	Regularisation	Lasso		5.87%	8.41%	7.35%	3.77%
15	Multiple Linear Regression	Regularisation	Ridge Regression		5.77%	8.61%	7.59%	4.98%
16	Multiple Linear Regression	Regularisation	Elastic Nets		5.87%	8.41%	7.35%	3.77%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	3.58%	5.08%	3.59%	5.32%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	3.87%	5.35%	3.49%	5.56%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	6.18%	8.88%	7.79%	4.39%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	3.81%	5.11%	3.31%	5.21%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	3.42%	4.99%	3.53%	4.96%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	3.81%	5.11%	3.31%	5.21%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	3.13%	3.55%	2.08%	2.66%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	67.25%	18.23%	2.00%	2.39%

Table H.25: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,6* (*cont.*)

	Method				e_{ma}					
					Test Set T_o, o=3	Test Set T_u, u=4	Test Set T_o, o=4	Test Set T_u, u=1		
Number (m)	Class I	Class II	Class III	Class IV	*I-II-A-5*	*I-II-A-6*	*I-II-A-7*	*I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			38.83%	3.02%	63.30%	81.92%	**29.23%**	**1.03**
2		Bootstrap Aggregating			153.29%	29.60%	15.10%	52.23%	**41.22%**	**1.14**
3	Neural Network	Feedforward			6.19%	2.19%	2.45%	4.31%	**3.49%**	**0.40**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.91%	9.74%	11.56%	2.31%	**7.63%**	**0.37**
7			Andrews		7.91%	9.74%	11.56%	2.31%	**7.63%**	**0.37**
8			Cauchy M-estimators by Moore		7.80%	9.71%	11.49%	2.33%	**7.60%**	**0.37**
9			Fair by Rey		3.69%	6.40%	5.00%	37.36%	**9.86%**	**1.14**
10			Huber		5.05%	8.09%	7.85%	22.19%	**8.79%**	**0.64**
11			Logistic Regression		4.86%	7.83%	7.43%	23.99%	**8.89%**	**0.71**
12			Hinch and Talwar		3.98%	6.18%	4.50%	45.11%	**10.62%**	**1.32**
13			Holland and Welsch		7.90%	9.74%	11.56%	2.31%	**7.63%**	**0.38**
14		Regularisation	Lasso		6.88%	3.97%	7.38%	84.62%	**16.03%**	**1.73**
15			Ridge Regression		9.99%	5.00%	8.09%	85.68%	**16.96%**	**1.64**
16			Elastic Nets		6.88%	3.97%	7.38%	84.62%	**16.03%**	**1.73**
17		Stepwise	Interaction	Bounded Steps	16.39%	6.87%	6.62%	8.97%	**7.05%**	**0.59**
18				Unbounded Steps	19.85%	6.71%	6.49%	11.93%	**7.91%**	**0.69**
19			Purequadratic	Bounded Steps	6.37%	3.47%	7.60%	87.31%	**16.50%**	**1.74**
20				Unbounded Steps	18.06%	6.41%	6.29%	10.82%	**7.38%**	**0.66**
21			Quadratic	Bounded Steps	13.17%	6.24%	5.98%	10.69%	**6.62%**	**0.53**
22				Unbounded Steps	18.06%	6.41%	6.29%	10.82%	**7.38%**	**0.66**
23			Polynomial	Bounded Steps	3.19%	2.92%	3.11%	4.58%	**3.15%**	**0.23**
24				Unbounded Steps	6.81%	1.65%	23.26%	4.67%	**15.78%**	**1.41**

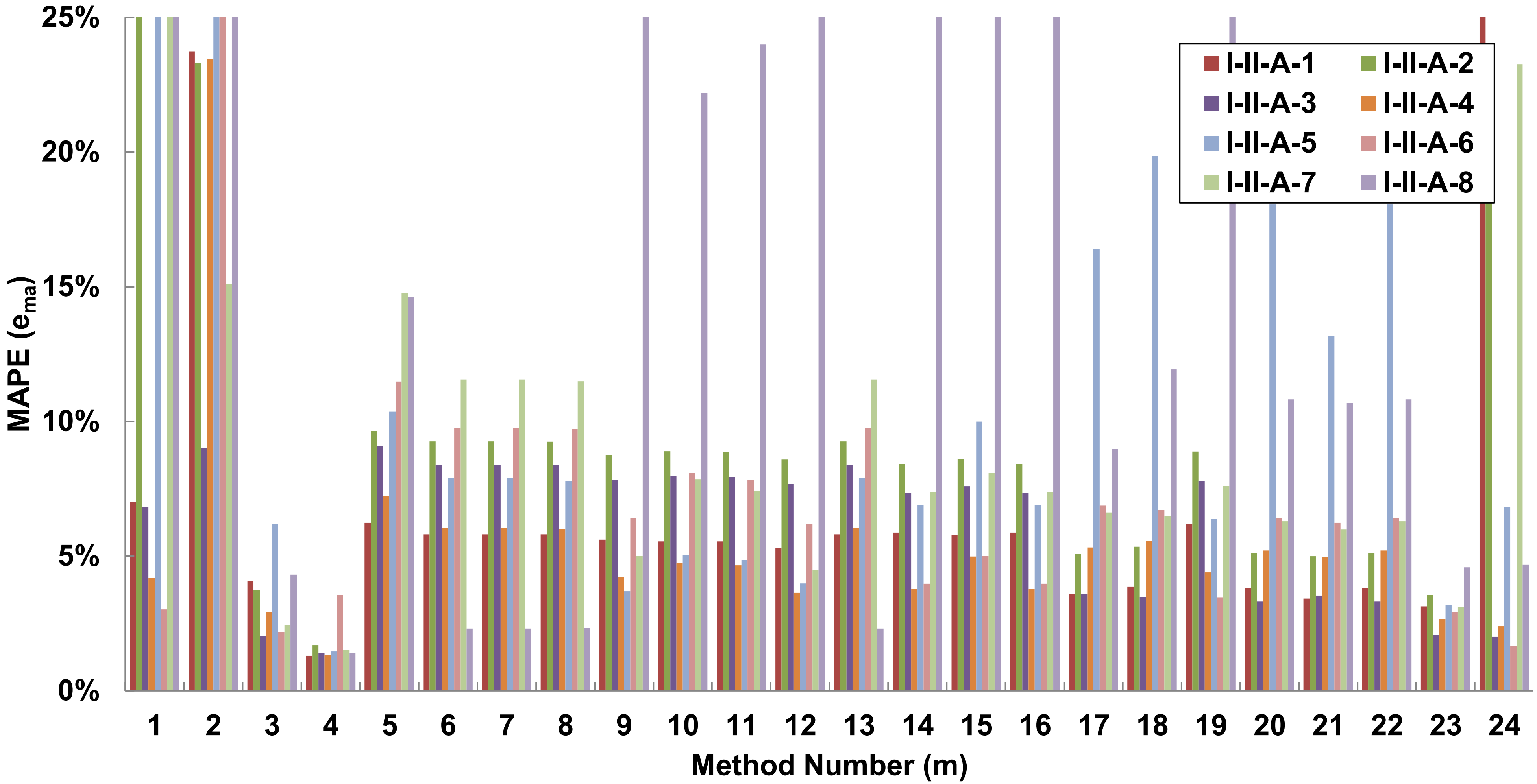


Figure H.25: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,6*

Table H.26: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,7*

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ *I-II-A-1*	e_{ma} Test Set $T_o, o=2$ *I-II-A-2*	e_{ma} Test Set $T_u, u=2$ *I-II-A-3*	e_{ma} Test Set $T_u, u=3$ *I-II-A-4*
1	Decision Tree	Boosting			7.10%	26.25%	6.97%	4.31%
2	Decision Tree	Bootstrap Aggregating			23.33%	22.83%	8.31%	21.27%
3	Neural Network	Feedforward			5.53%	5.11%	2.07%	3.11%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.34%	8.63%	7.74%	3.76%
7	Multiple Linear Regression	Robust	Andrews		5.34%	8.63%	7.74%	3.76%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		5.38%	8.68%	7.78%	3.87%
9	Multiple Linear Regression	Robust	Fair by Rey		5.50%	8.62%	7.66%	3.86%
10	Multiple Linear Regression	Robust	Huber		5.31%	8.55%	7.59%	3.74%
11	Multiple Linear Regression	Robust	Logistic Regression		5.32%	8.55%	7.59%	3.76%
12	Multiple Linear Regression	Robust	Hinch and Talwar		5.34%	8.64%	7.74%	3.77%
13	Multiple Linear Regression	Robust	Holland and Welsch		5.35%	8.65%	7.76%	3.81%
14	Multiple Linear Regression	Regularisation	Lasso		5.99%	8.58%	7.54%	3.79%
15	Multiple Linear Regression	Regularisation	Ridge Regression		5.90%	8.77%	7.77%	5.00%
16	Multiple Linear Regression	Regularisation	Elastic Nets		5.99%	8.58%	7.54%	3.79%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	3.42%	5.14%	3.60%	5.45%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	3.88%	5.35%	3.49%	5.61%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	6.30%	9.05%	7.99%	4.41%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	3.82%	5.16%	3.34%	5.26%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	3.43%	5.00%	3.53%	4.96%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	3.82%	5.16%	3.35%	5.26%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	2.91%	3.51%	2.09%	2.39%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	36.08%	15.01%	2.00%	2.51%

Table H.26: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,7* (*cont.*)

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_u, u=4 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			18.16%	17.75%	4.04%	71.95%	**19.57%**	**1.16**
2		Bootstrap Aggregating			135.58%	41.96%	14.17%	83.65%	**43.89%**	**1.00**
3	Neural Network	Feedforward			5.76%	1.98%	2.24%	5.50%	**3.91%**	**0.44**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		3.81%	6.42%	4.17%	41.48%	**10.17%**	**1.26**
7			Andrews		3.81%	6.42%	4.18%	41.48%	**10.17%**	**1.26**
8			Cauchy M-estimators by Moore		3.78%	6.64%	4.22%	38.46%	**9.85%**	**1.19**
9			Fair by Rey		3.40%	5.52%	3.06%	45.85%	**10.43%**	**1.38**
10			Huber		3.39%	5.69%	3.18%	45.82%	**10.41%**	**1.39**
11			Logistic Regression		3.38%	5.65%	3.15%	45.73%	**10.39%**	**1.39**
12			Hinch and Talwar		3.81%	6.43%	4.19%	41.32%	**10.16%**	**1.25**
13			Holland and Welsch		3.81%	6.52%	4.22%	40.28%	**10.05%**	**1.23**
14		Regularisation	Lasso		5.40%	3.90%	5.38%	73.73%	**14.29%**	**1.68**
15			Ridge Regression		9.80%	4.31%	5.73%	73.57%	**15.11%**	**1.57**
16			Elastic Nets		5.40%	3.90%	5.38%	73.73%	**14.29%**	**1.68**
17		Stepwise	Interaction	Bounded Steps	17.90%	5.90%	4.56%	21.93%	**8.49%**	**0.85**
18				Unbounded Steps	20.06%	5.89%	4.58%	21.78%	**8.83%**	**0.85**
19			Purequadratic	Bounded Steps	5.56%	3.44%	5.45%	74.80%	**14.63%**	**1.67**
20				Unbounded Steps	18.33%	5.53%	4.40%	21.54%	**8.42%**	**0.85**
21			Quadratic	Bounded Steps	13.20%	5.34%	4.21%	20.05%	**7.47%**	**0.80**
22				Unbounded Steps	18.26%	5.52%	4.39%	21.54%	**8.41%**	**0.85**
23			Polynomial	Bounded Steps	2.85%	2.04%	2.36%	8.09%	**3.28%**	**0.61**
24				Unbounded Steps	8.98%	14.76%	2.12%	3.06%	**10.57%**	**1.11**

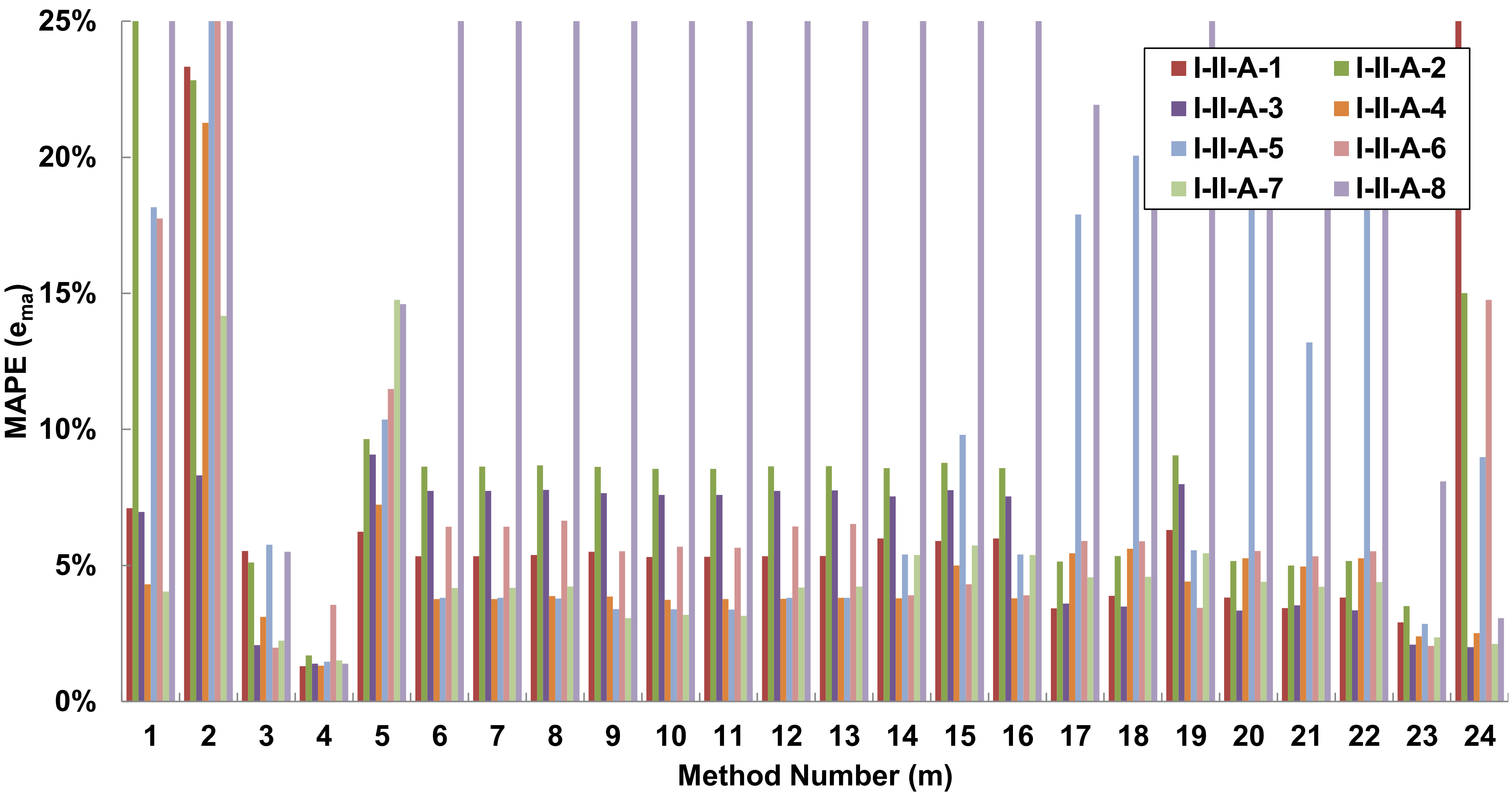


Figure H.26: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Training Set *I-II-A-8,3,4,7*

Table H.27: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu_{min}^{-1}}$

Number (m)	Class I	Class II	Class III	Class IV	$\mathbf{e_{ma}}$ Test Set $T_o, o=1$ *I-II-A-1*	Test Set $T_o, o=1$ *I-II-A-2*	Test Set $T_u, u=2$ *I-II-A-3*	Test Set $T_u, u=3$ *I-II-A-4*
1	Decision Tree	Boosting			6.97%	33.20%	6.77%	4.02%
2		Bootstrap Aggregating			15.60%	18.41%	3.53%	5.64%
3	Neural Network	Feedforward			4.61%	3.87%	2.16%	2.91%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.81%	9.26%	8.40%	6.05%
7			Andrews		5.81%	9.26%	8.40%	6.05%
8			Cauchy M-estimators by Moore		5.81%	9.25%	8.39%	5.99%
9			Fair by Rey		5.62%	8.81%	7.89%	4.43%
10			Huber		5.58%	8.98%	8.05%	5.03%
11			Logistic Regression		5.57%	8.95%	8.02%	4.95%
12			Hinch and Talwar		5.81%	9.26%	8.39%	6.05%
13			Holland and Welsch		5.81%	9.26%	8.40%	6.05%
14		Regularisation	Lasso		5.87%	8.42%	7.37%	3.77%
15			Ridge Regression		5.79%	8.65%	7.66%	4.89%
16			Elastic Nets		5.87%	8.42%	7.37%	3.77%
17		Stepwise	Interaction	Bounded Steps	3.65%	5.27%	3.28%	4.90%
18				Unbounded Steps	3.66%	5.54%	3.22%	4.93%
19			Purequadratic	Bounded Steps	6.17%	8.88%	7.80%	4.39%
20				Unbounded Steps	3.67%	5.16%	3.30%	4.81%
21			Quadratic	Bounded Steps	3.65%	5.44%	3.79%	4.21%
22				Unbounded Steps	3.66%	5.54%	3.22%	4.93%
23			Polynomial	Bounded Steps	34.98%	8.78%	1.89%	2.37%
24				Unbounded Steps				

Table H.27: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu_{\min}^{-1}}$ (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			40.95%	24.75%	104.70%	79.94%	**37.66%**	**0.98**
2		Bootstrap Aggregating			61.35%	78.81%	64.82%	3.86%	**31.50%**	**1.00**
3	Neural Network	Feedforward			5.12%	28.20%	30.48%	7.50%	**10.61%**	**1.10**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.89%	9.78%	11.59%	2.31%	**7.64%**	**0.38**
7			Andrews		7.89%	9.78%	11.59%	2.31%	**7.64%**	**0.38**
8			Cauchy M-estimators by Moore		7.79%	9.77%	11.55%	2.36%	**7.61%**	**0.37**
9			Fair by Rey		4.31%	8.03%	7.42%	32.80%	**9.91%**	**0.95**
10			Huber		5.85%	9.35%	9.89%	16.76%	**8.69%**	**0.43**
11			Logistic Regression		5.66%	9.20%	9.58%	18.61%	**8.82%**	**0.49**
12			Hinch and Talwar		7.89%	9.77%	11.58%	2.30%	**7.63%**	**0.38**
13			Holland and Welsch		7.88%	9.78%	11.59%	2.30%	**7.63%**	**0.38**
14		Regularisation	Lasso		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
15			Ridge Regression		9.21%	3.90%	7.85%	87.34%	**16.91%**	**1.69**
16			Elastic Nets		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
17		Stepwise	Interaction	Bounded Steps	15.04%	27.60%	26.27%	9.99%	**12.00%**	**0.83**
18				Unbounded Steps	15.63%	5.20%	7.55%	11.51%	**7.16%**	**0.60**
19			Purequadratic	Bounded Steps	6.32%	4.05%	7.98%	87.67%	**16.66%**	**1.73**
20				Unbounded Steps	14.66%	3.75%	6.62%	10.76%	**6.59%**	**0.62**
21			Quadratic	Bounded Steps	4.88%	5.49%	8.42%	39.95%	**9.48%**	**1.31**
22				Unbounded Steps	15.63%	5.20%	7.55%	11.51%	**7.16%**	**0.60**
23			Polynomial	Bounded Steps	5.15%	26.45%	37.56%	3.02%	**15.03%**	**1.02**
24				Unbounded Steps						

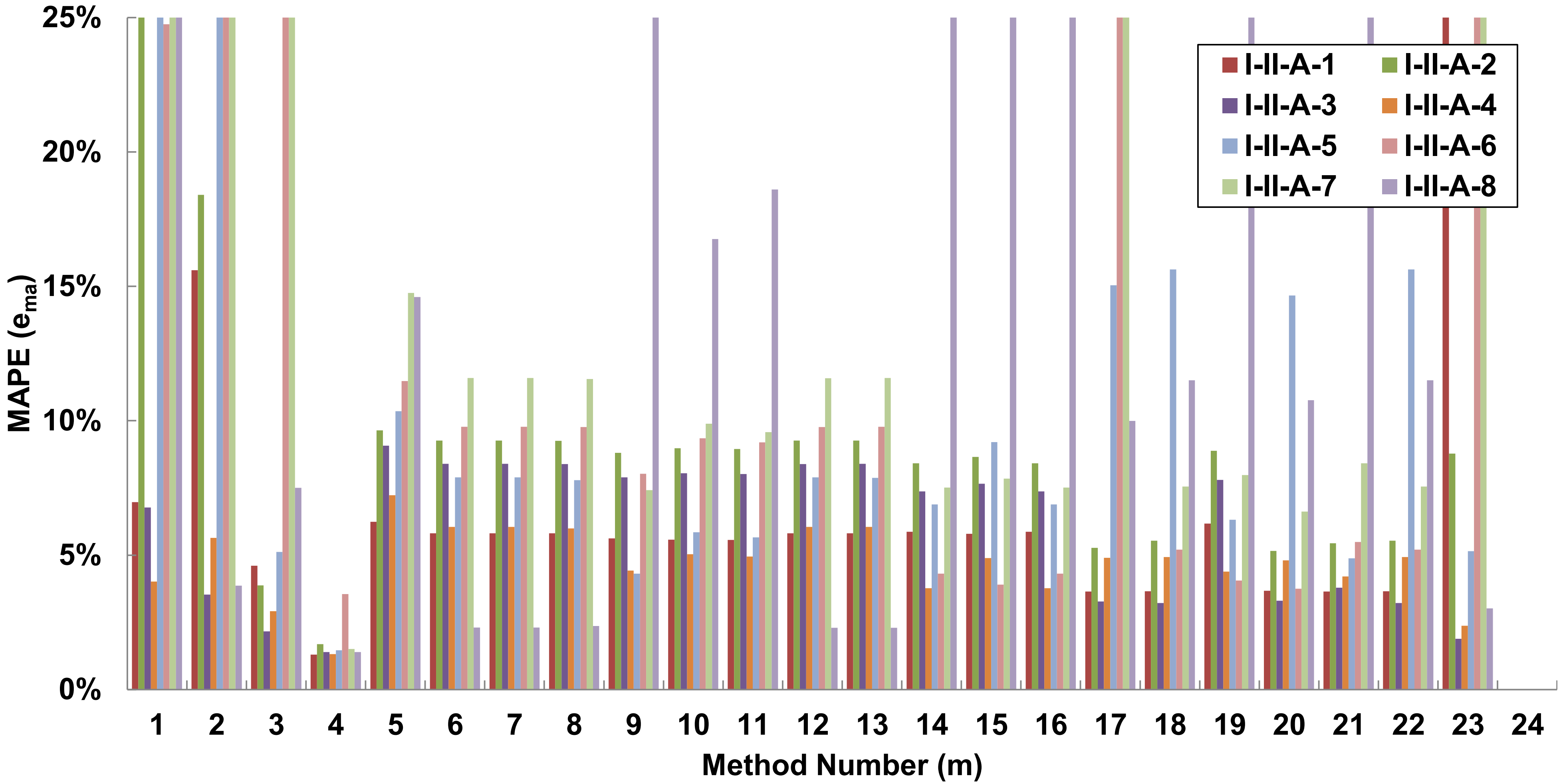


Figure H.27: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu_{\min}^{-1}}$

Table H.28: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor μ

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o,o=1 *I-II-A-1*	Test Set T_o,o=1 *I-II-A-2*	Test Set T_u,u=2 *I-II-A-3*	Test Set T_u,u=3 *I-II-A-4*
1	Decision Tree	Boosting			7.31%	73.08%	6.49%	3.76%
2		Bootstrap Aggregating			8.20%	7.61%	2.74%	3.93%
3	Neural Network	Feedforward			14.34%	7.47%	1.48%	1.57%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.81%	9.26%	8.40%	6.04%
7			Andrews		5.81%	9.26%	8.40%	6.04%
8			Cauchy M-estimators by Moore		5.82%	9.25%	8.39%	5.99%
9			Fair by Rey		5.64%	8.82%	7.88%	4.37%
10			Huber		5.58%	8.96%	8.04%	5.01%
11			Logistic Regression		5.57%	8.94%	8.01%	4.91%
12			Hinch and Talwar		5.81%	9.26%	8.40%	6.04%
13			Holland and Welsch		5.81%	9.26%	8.40%	6.04%
14		Regularisation	Lasso		5.87%	8.42%	7.37%	3.77%
15			Ridge Regression		5.77%	8.62%	7.61%	5.00%
16			Elastic Nets		5.87%	8.42%	7.37%	3.77%
17		Stepwise	Interaction	Bounded Steps	3.49%	4.80%	3.48%	4.44%
18				Unbounded Steps	3.71%	5.26%	3.34%	5.11%
19			Purequadratic	Bounded Steps	6.17%	8.88%	7.80%	4.39%
20				Unbounded Steps	3.64%	5.02%	3.31%	5.00%
21			Quadratic	Bounded Steps	3.28%	4.88%	3.48%	4.72%
22				Unbounded Steps	3.71%	5.17%	3.28%	4.98%
23			Polynomial	Bounded Steps	3.14%	3.56%	2.08%	2.69%
24				Unbounded Steps	53.23%	27.51%	16.01%	54.46%

Table H.28: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor μ (*cont.*)

	Method				$\mathbf{e_{ma}}$					
					Test Set T_o, o=3	**Test Set T_o, o=4**	**Test Set T_o, o=5**	**Test Set T_u, u=1**		
Number (m)	**Class I**	**Class II**	**Class III**	**Class IV**	***I-II-A-5***	***I-II-A-6***	***I-II-A-7***	***I-II-A-8***	μ_e	c_e
1	Decision Tree	Boosting			35.21%	19.51%	79.58%	73.86%	**37.35%**	**0.89**
2	Decision Tree	Bootstrap Aggregating			55.20%	68.16%	49.11%	3.61%	**24.82%**	**1.11**
3	Neural Network	Feedforward			7.41%	503.01%	734.64%	2.85%	**159.10%**	**1.83**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.86%	10.68%	13.00%	2.30%	**7.92%**	**0.41**
7	Multiple Linear Regression	Robust	Andrews		7.86%	10.68%	13.00%	2.30%	**7.92%**	**0.41**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		7.79%	10.69%	12.98%	2.30%	**7.90%**	**0.41**
9	Multiple Linear Regression	Robust	Fair by Rey		4.32%	12.88%	15.55%	34.46%	**11.74%**	**0.85**
10	Multiple Linear Regression	Robust	Huber		5.84%	11.53%	13.58%	18.10%	**9.58%**	**0.48**
11	Multiple Linear Regression	Robust	Logistic Regression		5.62%	11.43%	13.36%	19.68%	**9.69%**	**0.52**
12	Multiple Linear Regression	Robust	Hinch and Talwar		7.85%	10.72%	13.06%	2.30%	**7.93%**	**0.42**
13	Multiple Linear Regression	Robust	Holland and Welsch		7.86%	10.69%	13.01%	2.30%	**7.92%**	**0.41**
14	Multiple Linear Regression	Regularisation	Lasso		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
15	Multiple Linear Regression	Regularisation	Ridge Regression		10.05%	5.98%	8.91%	86.47%	**17.30%**	**1.62**
16	Multiple Linear Regression	Regularisation	Elastic Nets		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	11.92%	21.45%	48.15%	32.33%	**16.26%**	**1.02**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	16.93%	22.81%	23.72%	15.16%	**12.01%**	**0.72**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	6.32%	4.05%	7.98%	87.67%	**16.66%**	**1.73**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	16.55%	11.64%	22.26%	15.33%	**10.34%**	**0.69**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	11.64%	9.54%	20.78%	15.43%	**9.22%**	**0.69**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	16.22%	22.92%	21.93%	14.14%	**11.54%**	**0.72**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	3.23%	3.16%	3.45%	3.55%	**3.11%**	**0.16**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps	710.68%	2867.92%	6270.16%	193.91%	**1274.24%**	**1.76**

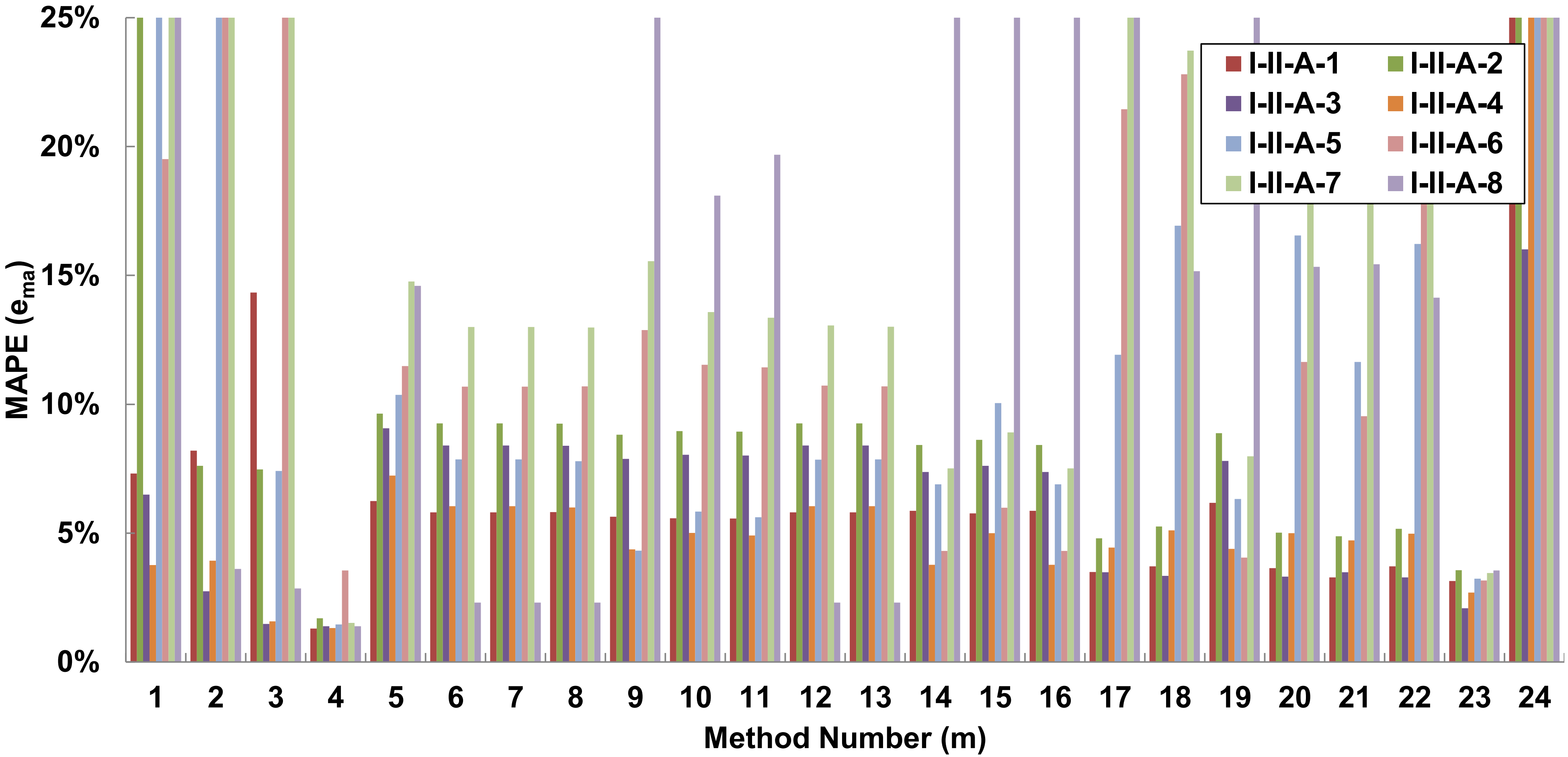


Figure H.28: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor μ

Table H.29: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor μ^{-1}

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1 I-II-A-1	Test Set T_o, o=1 I-II-A-2	Test Set T_u, u=2 I-II-A-3	Test Set T_u, u=3 I-II-A-4
1	Decision Tree	Boosting			6.86%	52.51%	6.49%	3.76%
2		Bootstrap Aggregating			8.94%	8.45%	3.31%	2.76%
3	Neural Network	Feedforward			5.16%	2.90%	1.66%	1.81%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.76%	9.23%	8.38%	6.04%
7			Andrews		5.76%	9.23%	8.38%	6.04%
8			Cauchy M-estimators by Moore		5.81%	9.25%	8.39%	6.00%
9			Fair by Rey		5.59%	8.73%	7.90%	4.39%
10			Huber		5.56%	8.92%	8.05%	4.98%
11			Logistic Regression		5.56%	8.89%	8.03%	4.90%
12			Hinch and Talwar		5.76%	9.23%	8.38%	6.04%
13			Holland and Welsch		5.77%	9.23%	8.38%	6.04%
14		Regularisation	Lasso		5.81%	8.50%	7.45%	4.02%
15			Ridge Regression		6.23%	9.30%	7.83%	4.53%
16			Elastic Nets		5.81%	8.50%	7.45%	4.02%
17		Stepwise	Interaction	Bounded Steps	4.11%	5.22%	3.15%	3.86%
18				Unbounded Steps	3.95%	5.08%	3.00%	3.82%
19			Purequadratic	Bounded Steps	6.27%	9.32%	7.81%	4.68%
20				Unbounded Steps	3.87%	5.01%	2.94%	3.79%
21			Quadratic	Bounded Steps	4.07%	5.18%	2.92%	3.68%
22				Unbounded Steps	4.05%	5.17%	2.91%	3.83%
23			Polynomial	Bounded Steps	2.24%	2.27%	1.58%	1.56%
24				Unbounded Steps				

Table H.29: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor μ^{-1} (*cont.*)

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma}: Test Set T_o,o=3 *I-II-A-5*	Test Set T_o,o=4 *I-II-A-6*	Test Set T_o,o=5 *I-II-A-7*	Test Set T_u,u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			35.21%	33.00%	102.44%	73.86%	**39.27%**	**0.90**
2		Bootstrap Aggregating			45.70%	60.63%	36.85%	3.64%	**21.29%**	**1.08**
3	Neural Network	Feedforward			3.31%	31.38%	27.10%	5.33%	**9.83%**	**1.23**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.88%	9.77%	11.58%	2.29%	**7.62%**	**0.38**
7			Andrews		7.88%	9.77%	11.58%	2.29%	**7.62%**	**0.38**
8			Cauchy M-estimators by Moore		7.79%	9.74%	11.53%	2.36%	**7.61%**	**0.37**
9			Fair by Rey		4.39%	7.83%	7.27%	32.60%	**9.84%**	**0.95**
10			Huber		5.78%	9.20%	9.76%	17.85%	**8.76%**	**0.47**
11			Logistic Regression		5.59%	8.98%	9.34%	19.20%	**8.81%**	**0.52**
12			Hinch and Talwar		7.86%	9.76%	11.58%	2.29%	**7.61%**	**0.38**
13			Holland and Welsch		7.87%	9.76%	11.58%	2.29%	**7.62%**	**0.38**
14		Regularisation	Lasso		6.70%	2.70%	6.01%	80.53%	**15.22%**	**1.74**
15			Ridge Regression		6.94%	3.09%	7.28%	86.92%	**16.52%**	**1.73**
16			Elastic Nets		6.70%	2.70%	6.01%	80.53%	**15.22%**	**1.74**
17		Stepwise	Interaction	Bounded Steps	4.45%	2.85%	3.14%	14.42%	**5.15%**	**0.74**
18				Unbounded Steps	8.13%	4.62%	5.12%	3.30%	**4.63%**	**0.35**
19			Purequadratic	Bounded Steps	7.54%	2.83%	6.78%	86.36%	**16.45%**	**1.72**
20				Unbounded Steps	8.02%	4.52%	5.05%	5.50%	**4.84%**	**0.32**
21			Quadratic	Bounded Steps	6.27%	39.96%	38.27%	3.19%	**12.94%**	**1.25**
22				Unbounded Steps	8.21%	30.58%	29.51%	5.35%	**11.20%**	**1.05**
23			Polynomial	Bounded Steps	2.10%	3.25%	3.15%	2.67%	**2.35%**	**0.27**
24				Unbounded Steps						

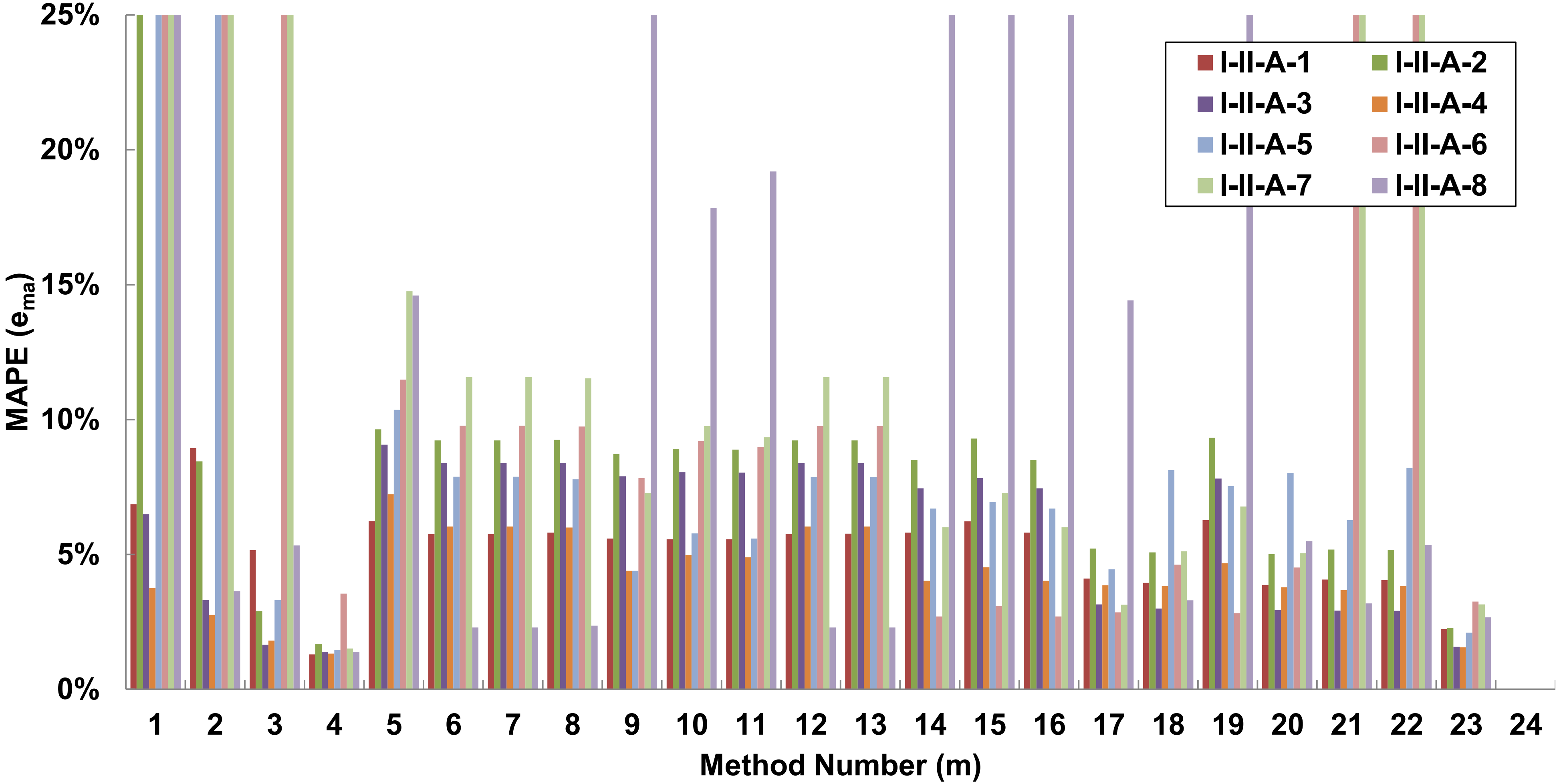


Figure H.29: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor μ^{-1}

Table H.30: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $\log \mu$

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma}: Test Set $T_o, o=1$ *I-II-A-1*	Test Set $T_o, o=1$ *I-II-A-2*	Test Set $T_u, u=2$ *I-II-A-3*	Test Set $T_u, u=3$ *I-II-A-4*
1	Decision Tree	Boosting			7.31%	69.93%	6.49%	3.76%
2		Bootstrap Aggregating			7.78%	7.44%	2.75%	3.12%
3	Neural Network	Feedforward			4.55%	3.30%	1.93%	2.67%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare			9.25%	8.39%	6.02%
7			Andrews		5.99%	9.25%	8.39%	6.02%
8			Cauchy M-estimators by Moore		3.48%	9.24%	8.39%	5.97%
9			Fair by Rey		5.79%	8.76%	7.83%	4.32%
10			Huber		5.79%	8.91%	8.00%	4.89%
11			Logistic Regression		5.80%	8.88%	7.96%	4.78%
12			Hinch and Talwar		5.59%	9.24%	8.37%	6.02%
13			Holland and Welsch		5.60%	9.25%	8.39%	6.01%
14		Regularisation	Lasso		5.60%	8.50%	7.45%	4.02%
15			Ridge Regression		5.78%	8.83%	7.83%	5.02%
16			Elastic Nets		5.79%	8.50%	7.45%	4.02%
17		Stepwise	Interaction	Bounded Steps	3.31%	4.40%	2.63%	4.75%
18				Unbounded Steps	3.46%	4.85%	2.57%	4.78%
19			Purequadratic	Bounded Steps	2.94%	8.88%	7.80%	4.39%
20				Unbounded Steps	3.43%	4.44%	2.65%	4.75%
21			Quadratic	Bounded Steps	3.31%	4.48%	2.87%	4.79%
22				Unbounded Steps	3.46%	4.58%	2.64%	4.79%
23			Polynomial	Bounded Steps	2.94%	3.34%	2.07%	2.81%
24				Unbounded Steps				

Table H.30: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $\log \mu$ (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma}: Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			35.21%	19.51%	79.58%	73.86%	**36.96%**	**0.88**
2		Bootstrap Aggregating			50.07%	62.37%	39.40%	3.58%	**22.06%**	**1.11**
3	Neural Network	Feedforward			7.86%	136.04%	103.47%	4.79%	**33.08%**	**1.64**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.85%	11.53%	13.95%	2.29%	**8.47%**	**0.44**
7			Andrews		7.85%	11.52%	13.94%	2.29%	**8.16%**	**0.44**
8			Cauchy M-estimators by Moore		7.79%	11.26%	13.55%	2.27%	**7.74%**	**0.49**
9			Fair by Rey		5.93%	18.52%	21.74%	34.21%	**13.39%**	**0.79**
10			Huber		6.09%	17.40%	20.94%	19.96%	**11.50%**	**0.59**
11			Logistic Regression		5.96%	18.26%	21.97%	22.30%	**11.99%**	**0.63**
12			Hinch and Talwar		7.84%	11.81%	14.35%	2.32%	**8.19%**	**0.46**
13			Holland and Welsch		7.84%	11.51%	13.92%	2.29%	**8.10%**	**0.45**
14		Regularisation	Lasso		6.70%	2.70%	6.01%	80.53%	**15.19%**	**1.74**
15			Ridge Regression		7.36%	12.46%	20.27%	87.85%	**19.43%**	**1.45**
16			Elastic Nets		6.70%	2.70%	6.01%	80.53%	**15.21%**	**1.74**
17		Stepwise	Interaction	Bounded Steps	6.67%	19.69%	34.58%	12.79%	**11.10%**	**1.00**
18				Unbounded Steps	8.29%	58.27%	57.95%	12.67%	**19.11%**	**1.27**
19			Purequadratic	Bounded Steps	6.32%	4.05%	57.95%	87.67%	**22.50%**	**1.43**
20				Unbounded Steps	6.29%	20.36%	34.39%	13.78%	**11.26%**	**0.99**
21			Quadratic	Bounded Steps	7.88%	9.06%	15.89%	10.06%	**7.29%**	**0.60**
22				Unbounded Steps	6.40%	47.26%	51.08%	13.19%	**16.68%**	**1.22**
23			Polynomial	Bounded Steps	3.43%	3.67%	3.93%	3.12%	**3.16%**	**0.18**
24				Unbounded Steps						

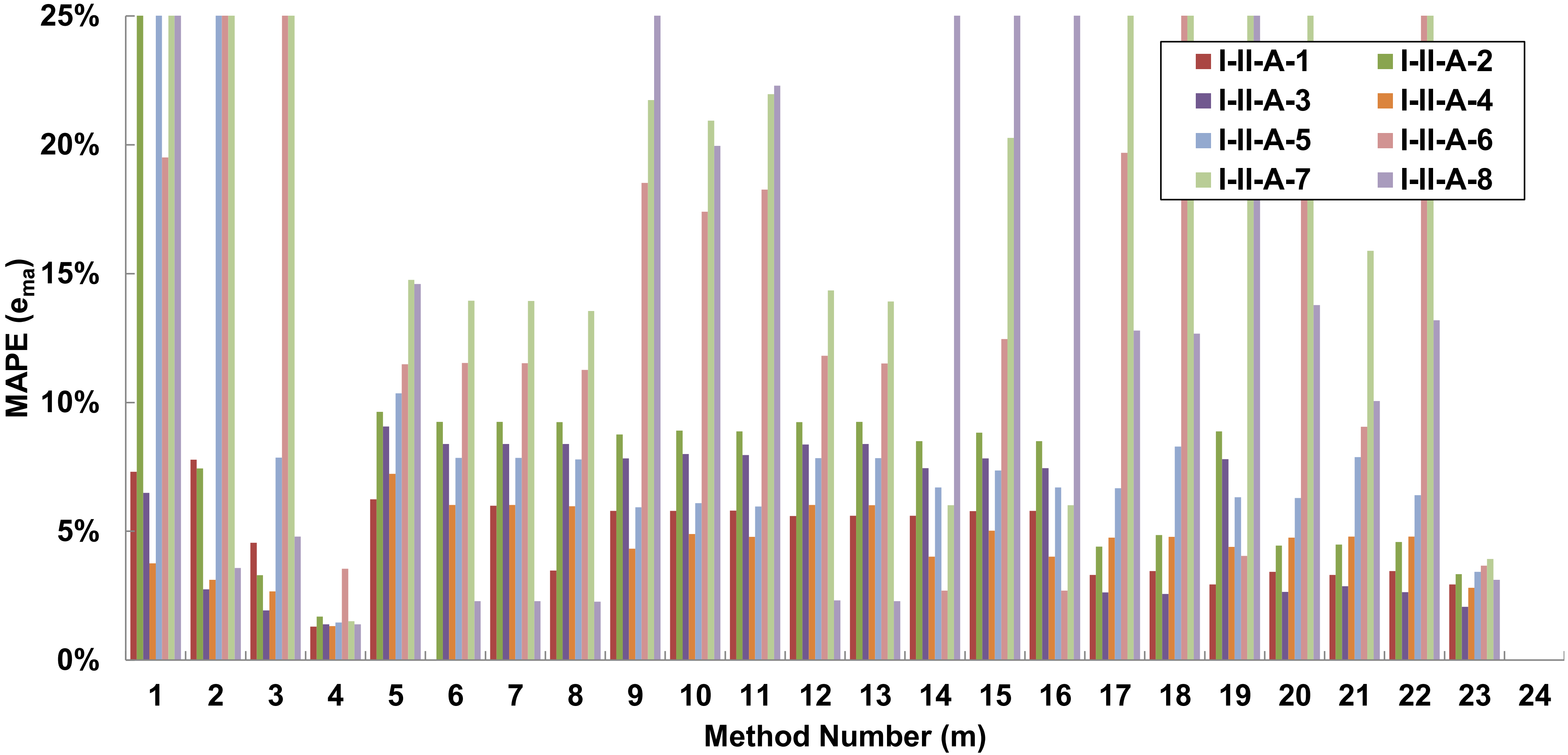


Figure H.30: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $\log \mu$

Table H.31: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1 I-II-A-1	Test Set T_o, o=1 I-II-A-2	Test Set T_u, u=2 I-II-A-3	Test Set T_u, u=3 I-II-A-4
1	Decision Tree	Boosting			6.86%	35.95%	6.49%	3.76%
2	Decision Tree	Bootstrap Aggregating			8.94%	8.45%	3.31%	2.76%
3	Neural Network	Feedforward			4.47%	2.69%	1.71%	1.99%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.76%	9.23%	8.37%	6.04%
7	Multiple Linear Regression	Robust	Andrews		5.76%	9.23%	8.37%	6.05%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		5.81%	9.25%	8.39%	6.00%
9	Multiple Linear Regression	Robust	Fair by Rey		5.63%	8.86%	7.90%	4.48%
10	Multiple Linear Regression	Robust	Huber		5.57%	8.96%	8.05%	5.02%
11	Multiple Linear Regression	Robust	Logistic Regression		5.57%	8.96%	8.04%	4.97%
12	Multiple Linear Regression	Robust	Hinch and Talwar		5.76%	9.23%	8.38%	6.05%
13	Multiple Linear Regression	Robust	Holland and Welsch		5.77%	9.24%	8.38%	6.04%
14	Multiple Linear Regression	Regularisation	Lasso		5.87%	8.42%	7.37%	3.77%
15	Multiple Linear Regression	Regularisation	Ridge Regression		5.91%	9.01%	7.61%	4.57%
16	Multiple Linear Regression	Regularisation	Elastic Nets		5.87%	8.42%	7.37%	3.77%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	4.35%	5.60%	3.37%	3.93%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	4.14%	5.18%	3.04%	4.03%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	6.37%	9.36%	7.84%	4.58%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	4.14%	5.29%	3.06%	3.84%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	4.40%	5.63%	3.38%	3.93%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	4.24%	5.40%	3.03%	4.02%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	1.48%	2.32%	1.81%	1.73%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps				

Table H.31: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$ (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o,o=3 *I-II-A-5*	Test Set T_o,o=4 *I-II-A-6*	Test Set T_o,o=5 *I-II-A-7*	Test Set T_u,u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			35.21%	33.00%	102.44%	73.86%	**37.20%**	**0.94**
2		Bootstrap Aggregating			45.70%	60.63%	36.85%	3.64%	**21.29%**	**1.08**
3	Neural Network	Feedforward			3.87%	9.44%	9.50%	4.17%	**4.73%**	**0.65**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.89%	9.73%	11.55%	2.29%	**7.61%**	**0.38**
7			Andrews		7.89%	9.73%	11.55%	2.29%	**7.61%**	**0.38**
8			Cauchy M-estimators by Moore		7.79%	9.74%	11.53%	2.36%	**7.61%**	**0.37**
9			Fair by Rey		4.35%	7.17%	6.68%	32.17%	**9.66%**	**0.96**
10			Huber		5.85%	9.06%	9.68%	17.39%	**8.70%**	**0.45**
11			Logistic Regression		5.73%	8.81%	9.25%	17.87%	**8.65%**	**0.47**
12			Hinch and Talwar		7.89%	9.73%	11.54%	2.30%	**7.61%**	**0.38**
13			Holland and Welsch		7.88%	9.74%	11.55%	2.29%	**7.61%**	**0.38**
14		Regularisation	Lasso		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
15			Ridge Regression		7.43%	4.75%	8.36%	86.05%	**16.71%**	**1.68**
16			Elastic Nets		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
17		Stepwise	Interaction	Bounded Steps	4.34%	3.53%	3.27%	17.50%	**5.74%**	**0.84**
18				Unbounded Steps	11.13%	6.39%	6.79%	2.60%	**5.41%**	**0.51**
19			Purequadratic	Bounded Steps	7.58%	2.79%	6.89%	86.49%	**16.49%**	**1.72**
20				Unbounded Steps	9.74%	7.37%	8.20%	5.61%	**5.91%**	**0.39**
21			Quadratic	Bounded Steps	4.34%	3.53%	3.27%	17.75%	**5.78%**	**0.85**
22				Unbounded Steps	10.94%	32.12%	31.36%	3.03%	**11.77%**	**1.07**
23			Polynomial	Bounded Steps	2.26%	2.31%	2.70%	2.34%	**2.12%**	**0.19**
24				Unbounded Steps						

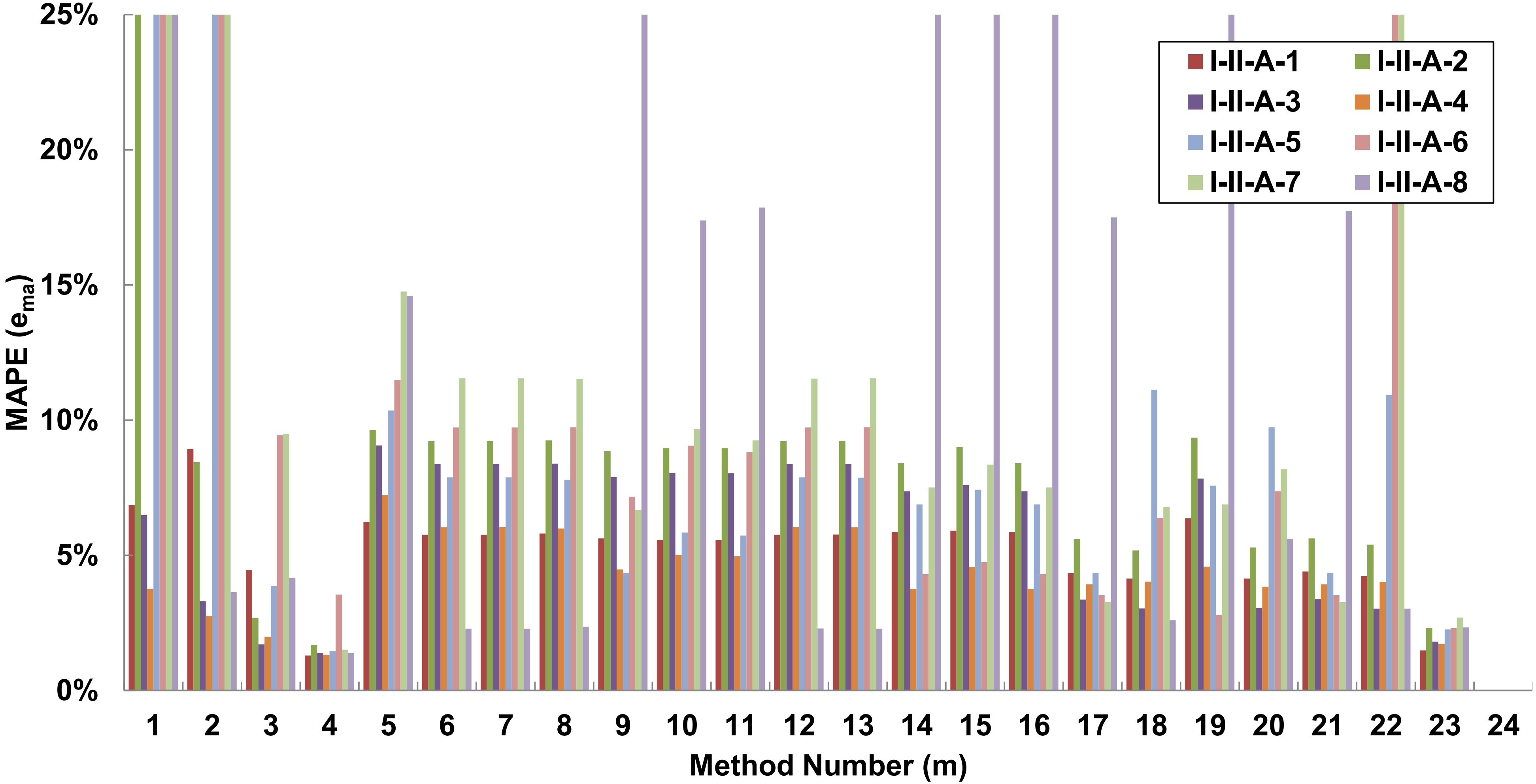


Figure H.31: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$

Table H.32: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor c

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1 I-II-A-1	Test Set T_o, o=1 I-II-A-2	Test Set T_u, u=2 I-II-A-3	Test Set T_u, u=3 I-II-A-4
1	Decision Tree	Boosting			6.82%	34.78%	6.24%	4.33%
2	Decision Tree	Bootstrap Aggregating			18.67%	17.39%	3.74%	9.67%
3	Neural Network	Feedforward			4.31%	3.73%	1.96%	3.40%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		84.61%	83.84%	82.75%	38.05%
7	Multiple Linear Regression	Robust	Andrews		84.60%	83.84%	82.75%	38.05%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		85.74%	83.47%	82.63%	22.00%
9	Multiple Linear Regression	Robust	Fair by Rey		84.56%	78.89%	76.47%	32.79%
10	Multiple Linear Regression	Robust	Huber		84.67%	81.24%	79.76%	26.47%
11	Multiple Linear Regression	Robust	Logistic Regression		84.59%	80.98%	79.42%	27.05%
12	Multiple Linear Regression	Robust	Hinch and Talwar		84.78%	83.90%	82.76%	38.19%
13	Multiple Linear Regression	Robust	Holland and Welsch		84.57%	83.82%	82.74%	38.01%
14	Multiple Linear Regression	Regularisation	Lasso		82.49%	81.60%	80.03%	48.04%
15	Multiple Linear Regression	Regularisation	Ridge Regression		65.55%	75.40%	77.46%	80.38%
16	Multiple Linear Regression	Regularisation	Elastic Nets		82.49%	81.60%	80.03%	48.04%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	3.61%	5.38%	3.58%	5.40%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	43.61%	17.92%	5.52%	12.87%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	6.17%	8.88%	7.80%	4.39%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	85.34%	61.92%	51.28%	109.63%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	3.50%	5.47%	4.19%	3.49%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	85.37%	61.92%	51.28%	109.50%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	3.14%	3.56%	2.08%	2.69%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps				

Table H.32: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor c (*cont.*)

Number (m)	Method: Class I	Class II	Class III	Class IV	$\mathbf{e_{ma}}$: Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			43.20%	23.83%	104.72%	76.33%	**37.53%**	**0.97**
2	Decision Tree	Bootstrap Aggregating			74.24%	75.47%	59.30%	3.78%	**32.78%**	**0.96**
3	Neural Network	Feedforward			5.17%	25.34%	24.07%	12.35%	**10.04%**	**0.95**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		61.16%	87.78%	79.92%	3.70%	**65.23%**	**0.46**
7	Multiple Linear Regression	Robust	Andrews		61.14%	87.78%	79.92%	3.69%	**65.22%**	**0.46**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		62.20%	87.78%	80.15%	6.36%	**63.79%**	**0.50**
9	Multiple Linear Regression	Robust	Fair by Rey		117.97%	87.79%	80.48%	12.41%	**71.42%**	**0.47**
10	Multiple Linear Regression	Robust	Huber		87.94%	87.74%	80.35%	11.70%	**67.48%**	**0.45**
11	Multiple Linear Regression	Robust	Logistic Regression		91.04%	87.75%	80.40%	12.24%	**67.93%**	**0.45**
12	Multiple Linear Regression	Robust	Hinch and Talwar		61.10%	87.77%	79.90%	3.63%	**65.25%**	**0.46**
13	Multiple Linear Regression	Robust	Holland and Welsch		61.25%	87.78%	79.94%	3.81%	**65.24%**	**0.46**
14	Multiple Linear Regression	Regularisation	Lasso		86.16%	30.44%	123.48%	1012.78%	**193.13%**	**1.72**
15	Multiple Linear Regression	Regularisation	Ridge Regression		109.44%	35.31%	24.63%	478.40%	**118.32%**	**1.25**
16	Multiple Linear Regression	Regularisation	Elastic Nets		86.16%	30.44%	123.48%	1012.78%	**193.13%**	**1.72**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	17.03%	38.48%	37.56%	2.88%	**14.24%**	**1.08**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	28.52%	462.65%	435.84%	13.40%	**127.54%**	**1.56**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	6.32%	4.05%	7.98%	87.67%	**16.66%**	**1.73**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	350.07%	87.25%	82.40%	89.28%	**114.65%**	**0.84**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	3.29%	2.69%	4.05%	44.95%	**8.95%**	**1.63**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	350.07%	87.02%	82.51%	88.34%	**114.50%**	**0.85**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	3.23%	3.16%	3.45%	3.55%	**3.11%**	**0.16**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps						

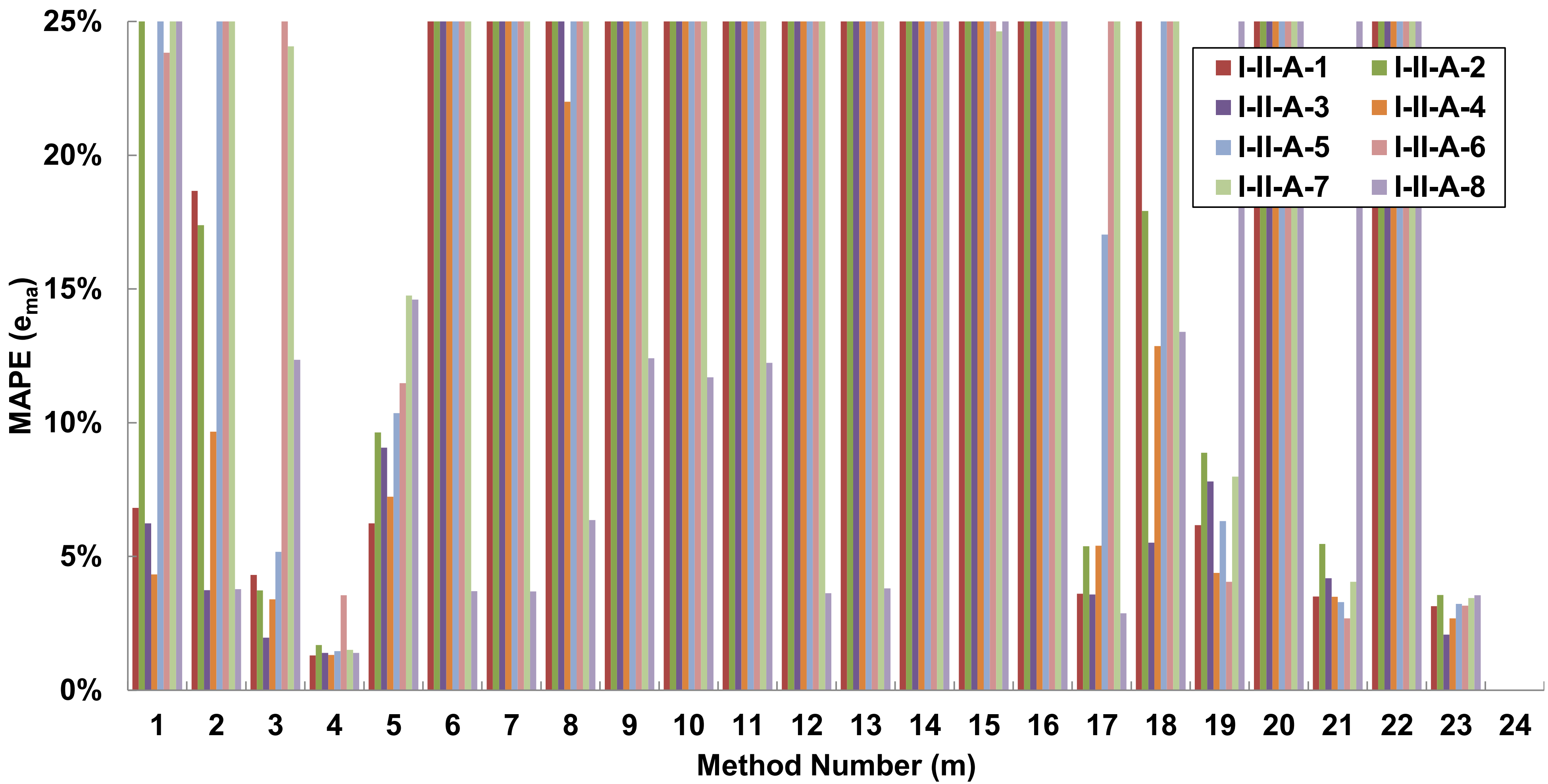


Figure H.32: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor c

Table H.33: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor c^{-1}

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma}: Test Set T_o, o=1 *I-II-A-1*	Test Set T_o, o=1 *I-II-A-2*	Test Set T_u, u=2 *I-II-A-3*	Test Set T_u, u=3 *I-II-A-4*
1	Decision Tree	Boosting			6.82%	35.75%	6.26%	4.29%
2		Bootstrap Aggregating			20.02%	17.17%	3.70%	9.60%
3	Neural Network	Feedforward			18.47%	4.21%	1.83%	2.35%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		81.99%	83.78%	82.75%	38.05%
7			Andrews		81.98%	83.78%	82.75%	38.05%
8			Cauchy M-estimators by Moore		85.56%	83.47%	82.63%	22.00%
9			Fair by Rey		83.52%	78.87%	76.47%	32.79%
10			Huber		83.88%	81.23%	79.76%	26.47%
11			Logistic Regression		83.80%	80.96%	79.42%	27.05%
12			Hinch and Talwar		82.14%	83.84%	82.76%	38.19%
13			Holland and Welsch		81.95%	83.76%	82.74%	38.01%
14		Regularisation	Lasso		79.44%	81.54%	80.03%	48.04%
15			Ridge Regression		65.98%	75.41%	77.46%	80.38%
16			Elastic Nets		79.44%	81.54%	80.03%	48.04%
17		Stepwise	Interaction	Bounded Steps	3.61%	5.38%	3.58%	5.40%
18				Unbounded Steps				
19			Purequadratic	Bounded Steps	6.17%	8.88%	7.80%	4.39%
20				Unbounded Steps	81.83%	61.88%	51.28%	109.63%
21			Quadratic	Bounded Steps	3.50%	5.47%	4.19%	3.49%
22				Unbounded Steps	81.81%	61.88%	51.28%	109.50%
23			Polynomial	Bounded Steps	3.14%	3.56%	2.08%	2.69%
24				Unbounded Steps				

Table H.33: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor c^{-1} (*cont.*)

	Method				$\mathbf{e_{ma}}$					
					Test Set T_o, o=3	**Test Set T_o, o=4**	**Test Set T_o, o=5**	**Test Set T_u, u=1**		
Number (m)	**Class I**	**Class II**	**Class III**	**Class IV**	***I-II-A-5***	***I-II-A-6***	***I-II-A-7***	***I-II-A-8***	μ_e	c_e
1	Decision Tree	Boosting			42.27%	26.68%	109.91%	78.20%	**38.77%**	**0.98**
2		Bootstrap Aggregating			77.10%	78.25%	63.88%	3.66%	**34.17%**	**0.97**
3	Neural Network	Feedforward			4.37%	68.58%	66.14%	3.02%	**21.12%**	**1.38**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		61.16%	87.78%	79.92%	3.70%	**64.89%**	**0.46**
7			Andrews		61.14%	87.78%	79.92%	3.69%	**64.89%**	**0.46**
8			Cauchy M-estimators by Moore		62.20%	87.78%	80.15%	6.36%	**63.77%**	**0.50**
9			Fair by Rey		117.97%	87.79%	80.48%	12.41%	**71.29%**	**0.47**
10			Huber		87.94%	87.74%	80.35%	11.70%	**67.38%**	**0.45**
11			Logistic Regression		91.04%	87.75%	80.40%	12.24%	**67.83%**	**0.45**
12			Hinch and Talwar		61.10%	87.77%	79.90%	3.63%	**64.92%**	**0.46**
13			Holland and Welsch		61.25%	87.78%	79.94%	3.81%	**64.91%**	**0.46**
14		Regularisation	Lasso		86.16%	30.44%	123.48%	1012.78%	**192.74%**	**1.73**
15			Ridge Regression		109.44%	35.31%	24.63%	478.40%	**118.38%**	**1.25**
16			Elastic Nets		86.16%	30.44%	123.48%	1012.78%	**192.74%**	**1.73**
17		Stepwise	Interaction	Bounded Steps	17.03%	38.48%	37.56%	2.88%	**14.24%**	**1.08**
18				Unbounded Steps						
19			Purequadratic	Bounded Steps	6.32%	4.05%	7.98%	87.67%	**16.66%**	**1.73**
20				Unbounded Steps	350.07%	87.25%	82.40%	89.28%	**114.20%**	**0.85**
21			Quadratic	Bounded Steps	3.29%	2.69%	4.05%	44.95%	**8.95%**	**1.63**
22				Unbounded Steps	350.07%	87.02%	82.51%	88.34%	**114.05%**	**0.85**
23			Polynomial	Bounded Steps	3.23%	3.16%	3.45%	3.55%	**3.11%**	**0.16**
24				Unbounded Steps						

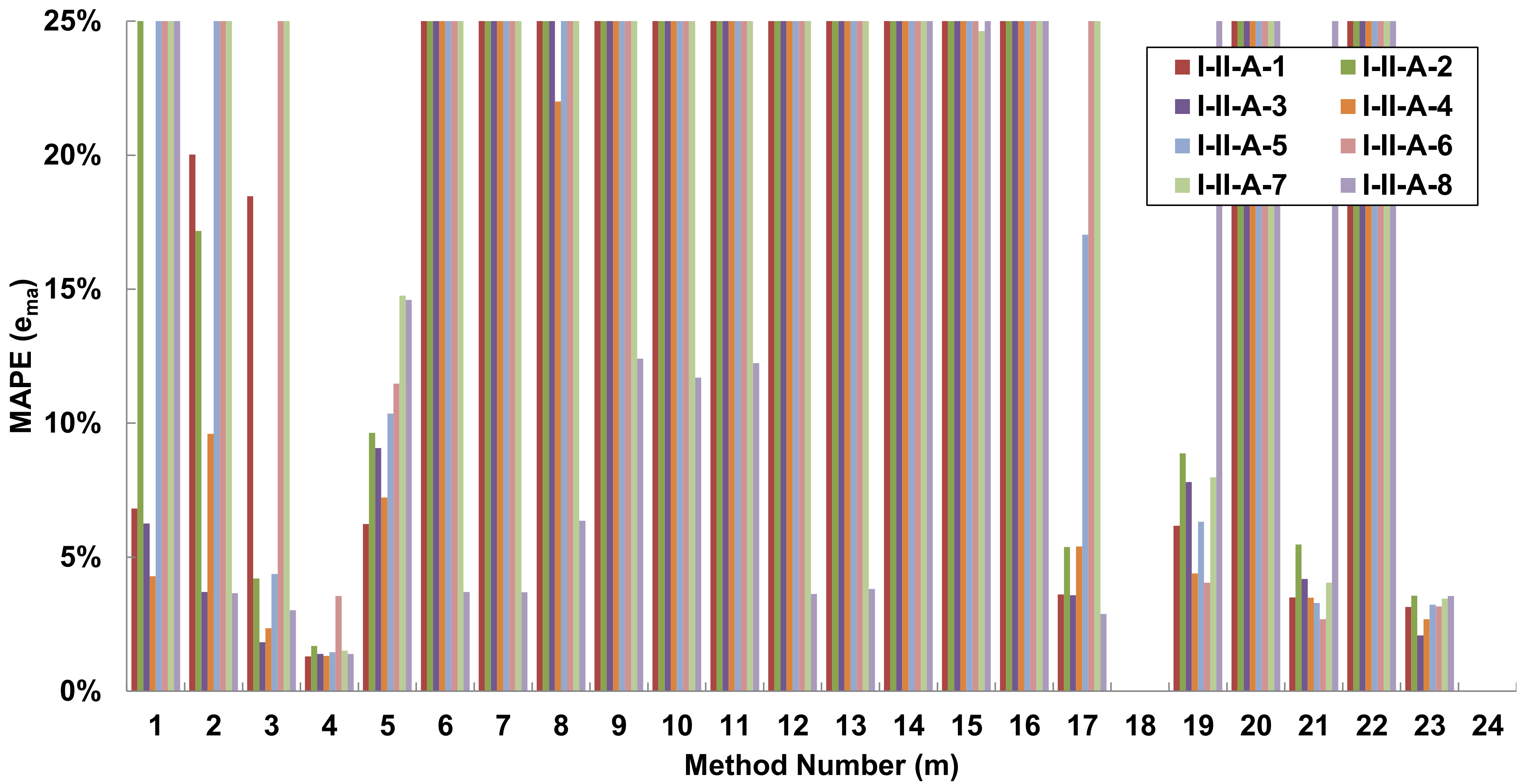


Figure H.33: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor c^{-1}

Table H.34: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $\log c$

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma}: Test Set T_o, o=1	Test Set T_o, o=1	Test Set T_u, u=2	Test Set T_u, u=3
					I-II-A-1	*I-II-A-2*	*I-II-A-3*	*I-II-A-4*
1	Decision Tree	Boosting			6.23%	35.45%	6.23%	4.41%
2		Bootstrap Aggregating			14.76%	12.69%	3.77%	9.72%
3	Neural Network	Feedforward			5.35%	4.47%	1.91%	2.39%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		84.20%	83.83%	82.75%	38.05%
7			Andrews		84.19%	83.83%	82.75%	38.05%
8			Cauchy M-estimators by Moore		85.52%	83.46%	82.63%	22.00%
9			Fair by Rey		84.28%	78.88%	76.47%	32.79%
10			Huber		84.40%	81.24%	79.76%	26.47%
11			Logistic Regression		84.32%	80.97%	79.42%	27.05%
12			Hinch and Talwar		84.38%	83.89%	82.76%	38.19%
13			Holland and Welsch		84.15%	83.81%	82.74%	38.01%
14		Regularisation	Lasso		82.02%	81.60%	80.03%	48.04%
15			Ridge Regression		64.95%	75.39%	77.46%	80.38%
16			Elastic Nets		82.02%	81.60%	80.03%	48.04%
17		Stepwise	Interaction	Bounded Steps	2.94%	5.36%	3.58%	5.40%
18				Unbounded Steps	43.34%	17.91%	5.52%	12.87%
19			Purequadratic	Bounded Steps	5.56%	8.86%	7.80%	4.39%
20				Unbounded Steps	84.89%	61.89%	51.28%	109.63%
21			Quadratic	Bounded Steps	2.78%	5.45%	4.19%	3.49%
22				Unbounded Steps	84.92%	61.90%	51.28%	109.50%
23			Polynomial	Bounded Steps	2.48%	3.54%	2.08%	2.69%
24				Unbounded Steps				

Table H.34: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $\log c$ (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=3$ *I-II-A-5*	Test Set $T_o, o=4$ *I-II-A-6*	Test Set $T_o, o=5$ *I-II-A-7*	Test Set $T_u, u=1$ *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			43.85%	24.17%	105.62%	76.02%	**37.75%**	**0.97**
2		Bootstrap Aggregating			77.09%	81.06%	68.55%	3.89%	**33.94%**	**1.03**
3	Neural Network	Feedforward			5.41%	69.37%	60.06%	2.99%	**18.99%**	**1.49**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		61.16%	87.78%	79.92%	3.70%	**65.17%**	**0.46**
7			Andrews		61.14%	87.78%	79.92%	3.69%	**65.17%**	**0.46**
8			Cauchy M-estimators by Moore		62.20%	87.78%	80.15%	6.36%	**63.76%**	**0.50**
9			Fair by Rey		117.97%	87.79%	80.48%	12.41%	**71.38%**	**0.47**
10			Huber		87.94%	87.74%	80.35%	11.70%	**67.45%**	**0.45**
11			Logistic Regression		91.04%	87.75%	80.40%	12.24%	**67.90%**	**0.45**
12			Hinch and Talwar		61.10%	87.77%	79.90%	3.63%	**65.20%**	**0.46**
13			Holland and Welsch		61.25%	87.78%	79.94%	3.81%	**65.19%**	**0.46**
14		Regularisation	Lasso		86.16%	30.44%	123.48%	1012.78%	**193.07%**	**1.72**
15			Ridge Regression		109.44%	35.31%	24.63%	478.40%	**118.25%**	**1.25**
16			Elastic Nets		86.16%	30.44%	123.48%	1012.78%	**193.07%**	**1.72**
17		Stepwise	Interaction	Bounded Steps	17.03%	38.48%	37.56%	2.88%	**14.15%**	**1.09**
18				Unbounded Steps	28.52%	462.65%	435.84%	13.40%	**127.51%**	**1.56**
19			Purequadratic	Bounded Steps	6.32%	4.05%	7.98%	87.67%	**16.58%**	**1.74**
20				Unbounded Steps	350.07%	87.25%	82.40%	89.28%	**114.59%**	**0.84**
21			Quadratic	Bounded Steps	3.29%	2.69%	4.05%	44.95%	**8.86%**	**1.65**
22				Unbounded Steps	350.07%	87.02%	82.51%	88.34%	**114.44%**	**0.85**
23			Polynomial	Bounded Steps	3.23%	3.16%	3.45%	3.55%	**3.02%**	**0.18**
24				Unbounded Steps						

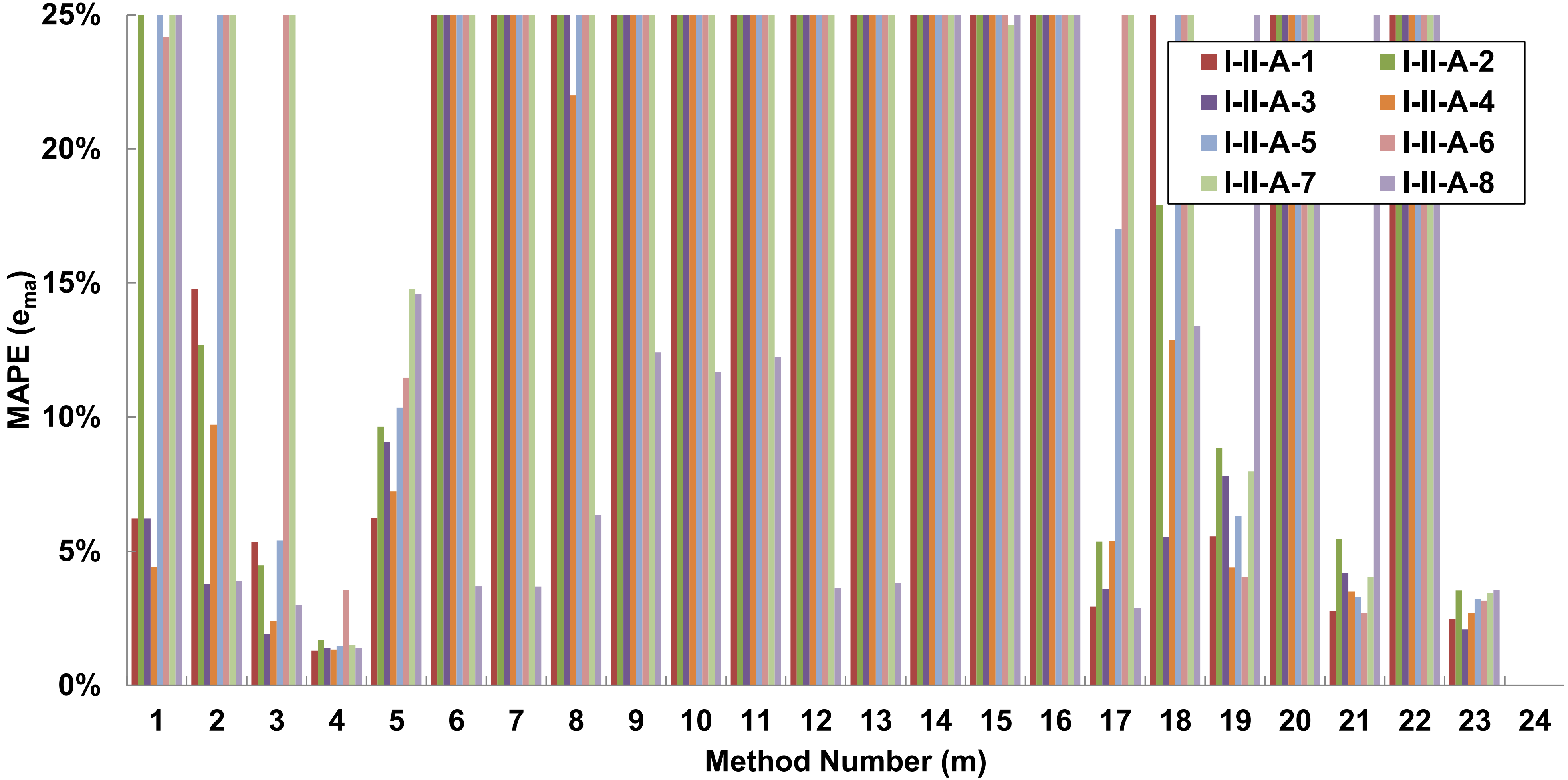


Figure H.34: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $\log c$

Table H.35: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, $e^{\mu_{\min}^{-1}}$

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ I-II-A-1	e_{ma} Test Set $T_o, o=1$ I-II-A-2	e_{ma} Test Set $T_u, u=2$ I-II-A-3	e_{ma} Test Set $T_u, u=3$ I-II-A-4
1	Decision Tree	Boosting			6.87%	38.17%	6.39%	3.79%
2	Decision Tree	Bootstrap Aggregating			9.05%	8.20%	3.19%	2.96%
3	Neural Network	Feedforward			7.53%	8.66%	5.59%	11.73%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.76%	9.23%	8.37%	6.04%
7	Multiple Linear Regression	Robust	Andrews		5.76%	9.23%	8.37%	6.05%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		5.81%	9.25%	8.39%	6.00%
9	Multiple Linear Regression	Robust	Fair by Rey		5.65%	8.89%	7.88%	4.48%
10	Multiple Linear Regression	Robust	Huber		5.58%	8.97%	8.05%	5.03%
11	Multiple Linear Regression	Robust	Logistic Regression		5.57%	8.95%	8.03%	4.95%
12	Multiple Linear Regression	Robust	Hinch and Talwar		5.76%	9.23%	8.37%	6.05%
13	Multiple Linear Regression	Robust	Holland and Welsch		5.77%	9.24%	8.38%	6.04%
14	Multiple Linear Regression	Regularisation	Lasso		5.87%	8.42%	7.37%	3.77%
15	Multiple Linear Regression	Regularisation	Ridge Regression		5.91%	9.02%	7.60%	4.57%
16	Multiple Linear Regression	Regularisation	Elastic Nets		5.87%	8.42%	7.37%	3.77%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	4.35%	5.60%	3.37%	3.93%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	4.30%	5.22%	3.04%	3.94%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	6.43%	9.44%	7.80%	4.54%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	4.14%	5.29%	3.06%	3.84%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	4.40%	5.63%	3.38%	3.93%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	4.38%	5.43%	3.03%	3.99%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	1.51%	2.33%	1.74%	1.74%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps				

Table H.35: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, $e^{\mu_{\min}^{-1}}$ (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o,o=3 *I-II-A-5*	Test Set T_o,o=4 *I-II-A-6*	Test Set T_o,o=5 *I-II-A-7*	Test Set T_u,u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			38.85%	34.65%	107.49%	74.58%	**38.85%**	**0.94**
2	Decision Tree	Bootstrap Aggregating			46.61%	66.17%	45.65%	3.62%	**23.18%**	**1.10**
3	Neural Network	Feedforward			32.43%	12.82%	16.62%	52.36%	**18.47%**	**0.87**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.89%	9.75%	11.56%	2.29%	**7.61%**	**0.38**
7	Multiple Linear Regression	Robust	Andrews		7.89%	9.75%	11.56%	2.29%	**7.61%**	**0.38**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		7.79%	9.77%	11.55%	2.36%	**7.62%**	**0.37**
9	Multiple Linear Regression	Robust	Fair by Rey		4.40%	8.10%	7.50%	32.45%	**9.92%**	**0.93**
10	Multiple Linear Regression	Robust	Huber		5.86%	9.38%	9.94%	16.93%	**8.72%**	**0.44**
11	Multiple Linear Regression	Robust	Logistic Regression		5.69%	9.21%	9.61%	18.43%	**8.81%**	**0.49**
12	Multiple Linear Regression	Robust	Hinch and Talwar		7.88%	9.74%	11.56%	2.29%	**7.61%**	**0.38**
13	Multiple Linear Regression	Robust	Holland and Welsch		7.88%	9.75%	11.56%	2.29%	**7.61%**	**0.38**
14	Multiple Linear Regression	Regularisation	Lasso		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
15	Multiple Linear Regression	Regularisation	Ridge Regression		7.49%	5.24%	8.59%	85.90%	**16.79%**	**1.67**
16	Multiple Linear Regression	Regularisation	Elastic Nets		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	4.34%	3.53%	3.27%	17.50%	**5.74%**	**0.84**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	10.61%	16.81%	17.04%	3.53%	**8.06%**	**0.74**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	7.49%	6.11%	8.97%	86.25%	**17.13%**	**1.63**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	9.74%	7.37%	8.20%	5.61%	**5.91%**	**0.39**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	4.34%	3.53%	3.27%	17.75%	**5.78%**	**0.85**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	11.23%	41.46%	40.34%	6.06%	**14.49%**	**1.14**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	2.53%	2.44%	3.02%	2.90%	**2.28%**	**0.25**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps						

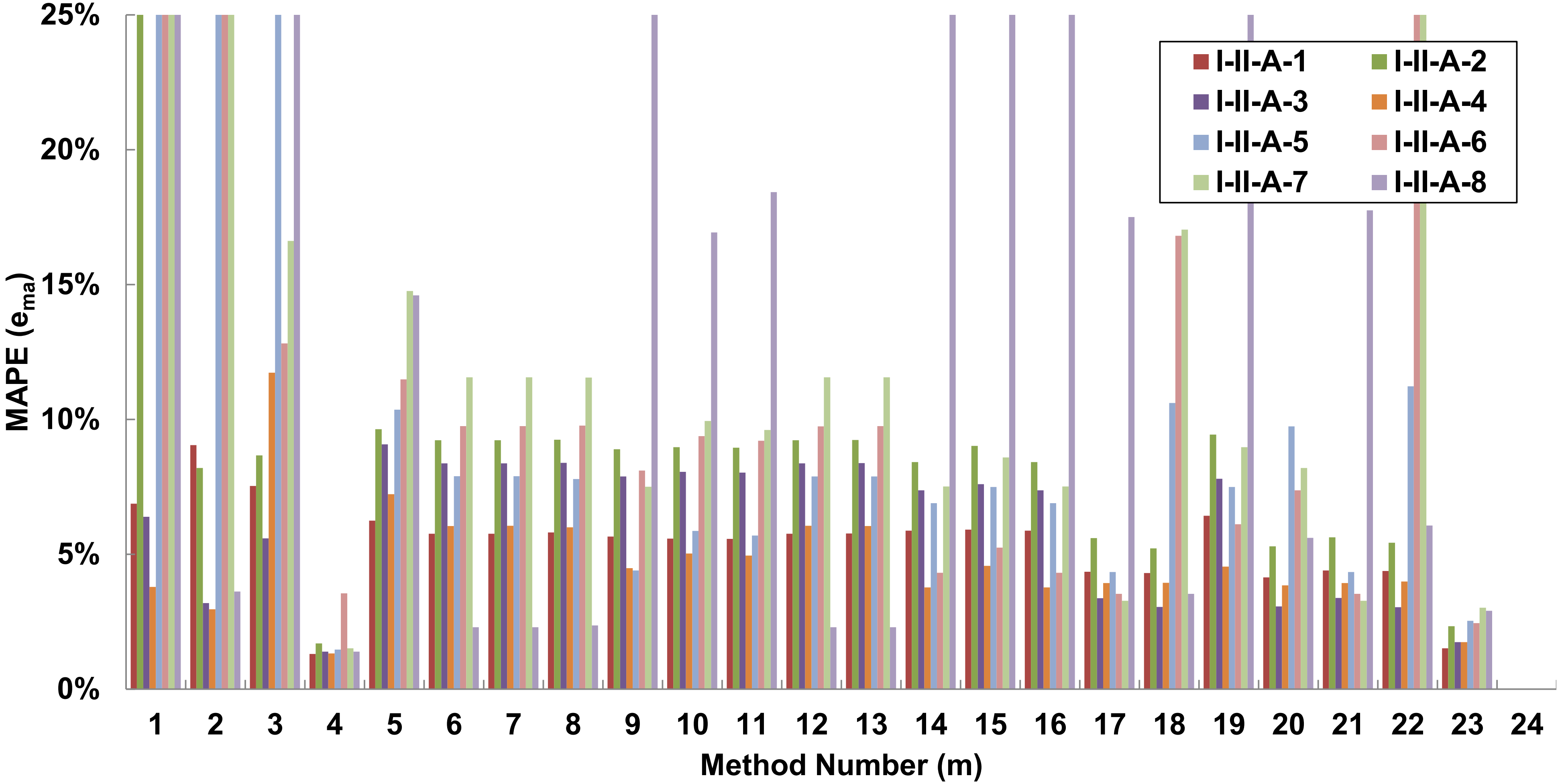


Figure H.35: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}, e^{\mu_{\min}^{-1}}$

Table H.36: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, μ

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=1 *I-II-A-1*	Test Set T_o, o=1 *I-II-A-2*	Test Set T_u, u=2 *I-II-A-3*	Test Set T_u, u=3 *I-II-A-4*
1	Decision Tree	Boosting			7.37%	74.62%	6.49%	3.76%
2	Decision Tree	Bootstrap Aggregating			9.00%	8.74%	3.49%	3.02%
3	Neural Network	Feedforward			3.97%	2.65%	1.62%	1.93%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.77%	9.23%	8.38%	6.04%
7	Multiple Linear Regression	Robust	Andrews		5.77%	9.23%	8.38%	6.04%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		5.82%	9.25%	8.39%	5.99%
9	Multiple Linear Regression	Robust	Fair by Rey		5.66%	8.88%	7.88%	4.41%
10	Multiple Linear Regression	Robust	Huber		5.57%	8.95%	8.04%	5.00%
11	Multiple Linear Regression	Robust	Logistic Regression		5.57%	8.93%	8.01%	4.91%
12	Multiple Linear Regression	Robust	Hinch and Talwar		5.77%	9.24%	8.38%	6.04%
13	Multiple Linear Regression	Robust	Holland and Welsch		5.77%	9.24%	8.38%	6.04%
14	Multiple Linear Regression	Regularisation	Lasso		5.87%	8.42%	7.37%	3.77%
15	Multiple Linear Regression	Regularisation	Ridge Regression		5.91%	9.01%	7.61%	4.57%
16	Multiple Linear Regression	Regularisation	Elastic Nets		5.87%	8.42%	7.37%	3.77%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	4.35%	5.60%	3.37%	3.93%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	4.17%	5.21%	3.00%	3.99%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	6.37%	9.36%	7.84%	4.58%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	4.02%	5.19%	3.03%	3.91%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	4.40%	5.63%	3.38%	3.93%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	4.26%	5.41%	3.02%	4.02%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	1.48%	2.32%	1.81%	1.73%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps				

Table H.36: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, μ (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			35.21%	16.98%	75.28%	73.86%	**36.70%**	**0.90**
2	Decision Tree	Bootstrap Aggregating			44.64%	52.29%	25.28%	3.67%	**18.77%**	**1.05**
3	Neural Network	Feedforward			5.36%	271.77%	320.85%	5.44%	**76.70%**	**1.78**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.85%	10.61%	12.90%	2.30%	**7.89%**	**0.41**
7	Multiple Linear Regression	Robust	Andrews		7.85%	10.61%	12.89%	2.30%	**7.88%**	**0.41**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		7.79%	10.69%	12.99%	2.30%	**7.90%**	**0.41**
9	Multiple Linear Regression	Robust	Fair by Rey		4.36%	12.91%	15.74%	33.94%	**11.72%**	**0.84**
10	Multiple Linear Regression	Robust	Huber		5.84%	11.48%	13.45%	18.22%	**9.57%**	**0.48**
11	Multiple Linear Regression	Robust	Logistic Regression		5.63%	11.38%	13.26%	19.62%	**9.66%**	**0.52**
12	Multiple Linear Regression	Robust	Hinch and Talwar		7.84%	10.69%	13.02%	2.31%	**7.91%**	**0.41**
13	Multiple Linear Regression	Robust	Holland and Welsch		7.85%	10.64%	12.93%	2.29%	**7.89%**	**0.41**
14	Multiple Linear Regression	Regularisation	Lasso		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
15	Multiple Linear Regression	Regularisation	Ridge Regression		7.43%	4.75%	8.36%	86.05%	**16.71%**	**1.68**
16	Multiple Linear Regression	Regularisation	Elastic Nets		6.89%	4.31%	7.51%	84.10%	**16.03%**	**1.72**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	4.34%	3.53%	3.27%	17.50%	**5.74%**	**0.84**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	10.68%	8.01%	8.94%	2.85%	**5.86%**	**0.51**
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	7.58%	2.79%	6.89%	86.49%	**16.49%**	**1.72**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	7.37%	7.20%	12.97%	12.35%	**7.01%**	**0.54**
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	4.34%	3.53%	3.27%	17.75%	**5.78%**	**0.85**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	9.83%	11.65%	17.15%	8.76%	**8.01%**	**0.60**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	2.26%	2.31%	2.70%	2.34%	**2.12%**	**0.19**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps						

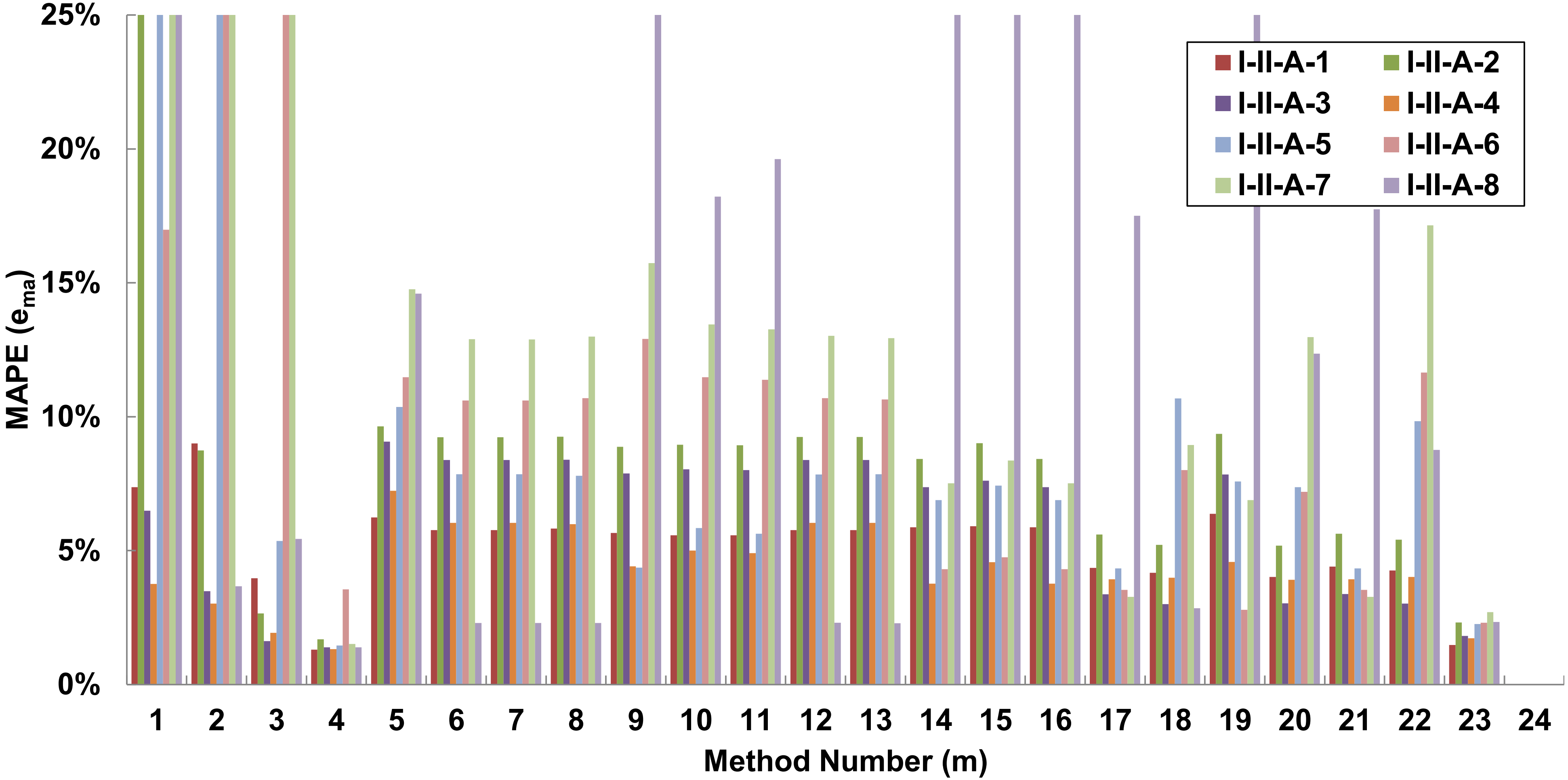


Figure H.36: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, μ

Table H.37: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, μ^{-1}

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma}: Test Set T_o, o=1 *I-II-A-1*	Test Set T_o, o=1 *I-II-A-2*	Test Set T_u, u=2 *I-II-A-3*	Test Set T_u, u=3 *I-II-A-4*
1	Decision Tree	Boosting			6.86%	52.51%	6.49%	3.76%
2		Bootstrap Aggregating			9.15%	9.09%	3.59%	2.91%
3	Neural Network	Feedforward			11.57%	9.37%	1.48%	1.81%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.90%	9.30%	8.46%	6.01%
7			Andrews		6.05%	9.39%	8.25%	6.03%
8			Cauchy M-estimators by Moore		5.83%	9.26%	8.41%	5.99%
9			Fair by Rey		5.61%	8.86%	7.82%	4.47%
10			Huber		5.62%	8.99%	8.04%	5.02%
11			Logistic Regression		5.63%	8.99%	8.02%	4.97%
12			Hinch and Talwar		5.78%	9.24%	8.37%	6.02%
13			Holland and Welsch		5.89%	9.30%	8.45%	6.01%
14		Regularisation	Lasso		5.81%	8.50%	7.45%	4.02%
15			Ridge Regression		6.25%	9.28%	7.85%	4.54%
16			Elastic Nets		5.81%	8.50%	7.45%	4.02%
17		Stepwise	Interaction	Bounded Steps	4.11%	5.22%	3.15%	3.86%
18				Unbounded Steps	4.52%	4.87%	2.34%	3.86%
19			Purequadratic	Bounded Steps	6.27%	9.32%	7.81%	4.68%
20				Unbounded Steps	4.37%	4.36%	2.36%	3.80%
21			Quadratic	Bounded Steps	4.34%	4.32%	2.36%	3.79%
22				Unbounded Steps	4.52%	4.89%	2.33%	3.82%
23			Polynomial	Bounded Steps	16.38%	4.98%	1.45%	1.42%
24				Unbounded Steps				

Table H.37: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, μ^{-1} (cont.)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			35.21%	33.00%	102.44%	73.86%	**39.27%**	**0.90**
2		Bootstrap Aggregating			44.82%	55.11%	29.86%	3.70%	**19.78%**	**1.05**
3	Neural Network	Feedforward			8.59%	29.42%	35.24%	3.43%	**12.61%**	**1.01**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.87%	10.44%	12.21%	2.30%	**7.81%**	**0.39**
7			Andrews		7.86%	11.00%	12.83%	2.28%	**7.96%**	**0.41**
8			Cauchy M-estimators by Moore		7.80%	9.84%	11.62%	2.35%	**7.64%**	**0.38**
9			Fair by Rey		5.16%	9.20%	8.71%	31.03%	**10.11%**	**0.86**
10			Huber		6.06%	10.01%	10.60%	16.59%	**8.87%**	**0.42**
11			Logistic Regression		5.98%	9.94%	10.38%	17.07%	**8.87%**	**0.44**
12			Hinch and Talwar		7.84%	10.11%	11.88%	2.28%	**7.69%**	**0.39**
13			Holland and Welsch		7.86%	10.40%	12.18%	2.30%	**7.80%**	**0.39**
14		Regularisation	Lasso		6.70%	2.70%	6.01%	80.53%	**15.22%**	**1.74**
15			Ridge Regression		6.96%	3.01%	7.23%	87.07%	**16.52%**	**1.73**
16			Elastic Nets		6.70%	2.70%	6.01%	80.53%	**15.22%**	**1.74**
17		Stepwise	Interaction	Bounded Steps	4.45%	2.85%	3.14%	14.42%	**5.15%**	**0.74**
18				Unbounded Steps	5.56%	103.03%	97.72%	4.33%	**28.28%**	**1.57**
19			Purequadratic	Bounded Steps	7.54%	2.83%	6.78%	86.36%	**16.45%**	**1.72**
20				Unbounded Steps	4.60%	15.93%	15.85%	6.01%	**7.16%**	**0.77**
21			Quadratic	Bounded Steps	4.18%	14.34%	14.13%	3.15%	**6.33%**	**0.78**
22				Unbounded Steps	5.91%	104.15%	98.88%	6.25%	**28.84%**	**1.56**
23			Polynomial	Bounded Steps	3.41%	4.88%	6.35%	2.46%	**5.17%**	**0.94**
24				Unbounded Steps						

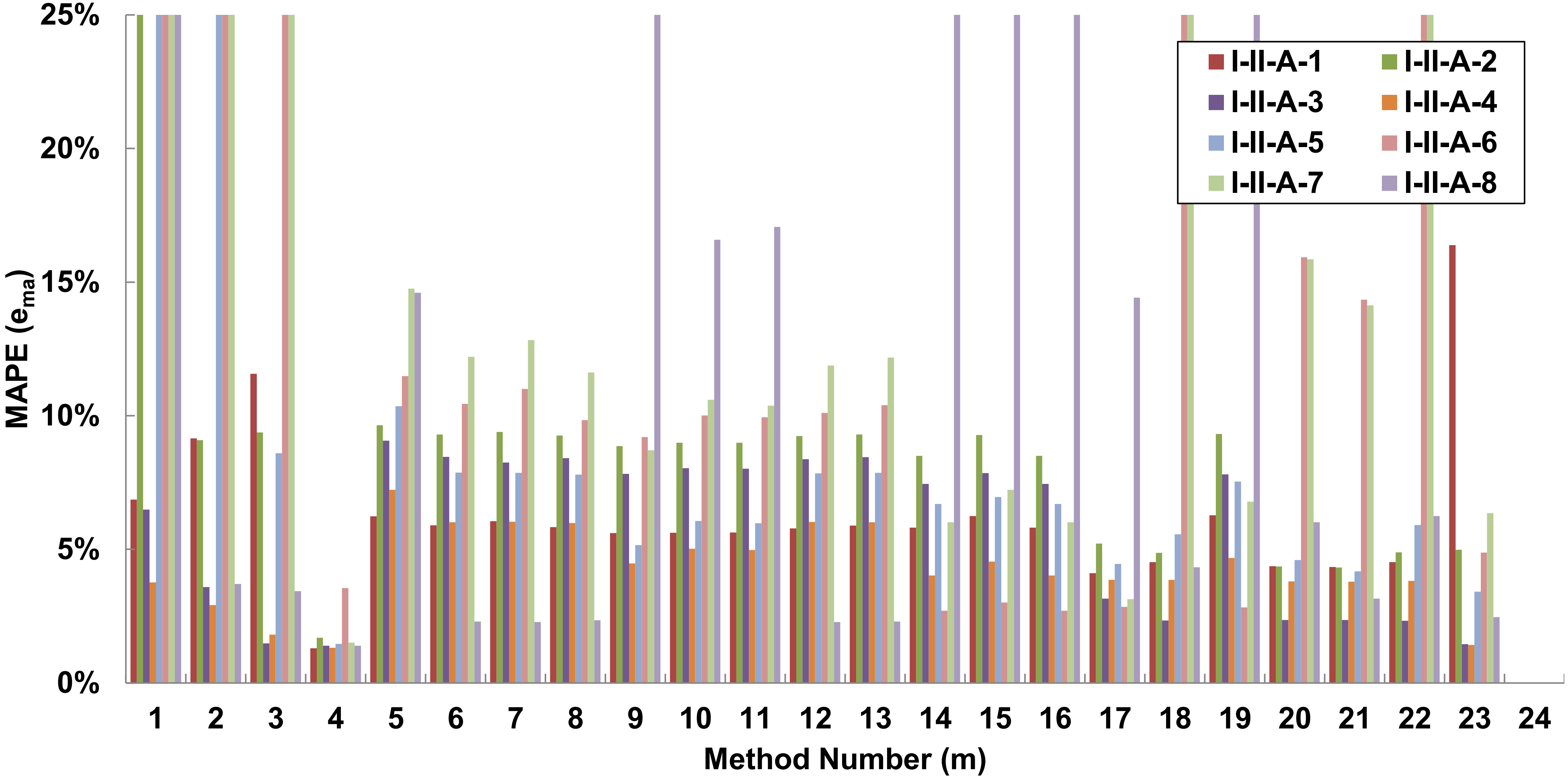


Figure H.37: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}, \mu^{-1}$

Table H.38: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, $\log\mu$

Number (m)	Method: Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ I-II-A-1	Test Set $T_o, o=1$ I-II-A-2	Test Set $T_u, u=2$ I-II-A-3	Test Set $T_u, u=3$ I-II-A-4
1	Decision Tree	Boosting			7.37%	71.46%	6.49%	3.76%
2	Decision Tree	Bootstrap Aggregating			9.00%	8.74%	3.49%	3.01%
3	Neural Network	Feedforward			2.65%	2.66%	1.53%	1.43%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		5.78%	9.24%	8.38%	6.02%
7	Multiple Linear Regression	Robust	Andrews		5.78%	9.24%	8.38%	6.02%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		5.83%	9.26%	8.40%	5.97%
9	Multiple Linear Regression	Robust	Fair by Rey		5.61%	8.86%	7.82%	4.37%
10	Multiple Linear Regression	Robust	Huber		5.62%	8.95%	7.99%	4.92%
11	Multiple Linear Regression	Robust	Logistic Regression		5.63%	8.94%	7.96%	4.81%
12	Multiple Linear Regression	Robust	Hinch and Talwar		5.82%	9.26%	8.41%	6.02%
13	Multiple Linear Regression	Robust	Holland and Welsch		5.78%	9.24%	8.38%	6.01%
14	Multiple Linear Regression	Regularisation	Lasso		5.81%	8.50%	7.45%	4.02%
15	Multiple Linear Regression	Regularisation	Ridge Regression		6.04%	9.08%	7.77%	4.84%
16	Multiple Linear Regression	Regularisation	Elastic Nets		5.81%	8.50%	7.45%	4.02%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	4.49%	4.77%	2.89%	4.57%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	3.80%	4.69%	2.68%	4.51%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	6.37%	9.36%	7.84%	4.58%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	3.78%	4.87%	2.72%	4.51%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	4.40%	5.63%	3.38%	3.93%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	4.18%	4.88%	2.72%	4.49%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	1.48%	2.32%	1.81%	1.73%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps				

Table H.38: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, $\log \mu$ (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			35.21%	16.98%	75.28%	73.86%	**36.30%**	**0.89**
2		Bootstrap Aggregating			44.64%	52.29%	25.28%	3.67%	**18.77%**	**1.05**
3	Neural Network	Feedforward			5.10%	67.10%	62.91%	2.62%	**18.25%**	**1.58**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.84%	11.40%	13.78%	2.28%	**8.09%**	**0.44**
7			Andrews		7.84%	11.39%	13.77%	2.28%	**8.09%**	**0.44**
8			Cauchy M-estimators by Moore		7.80%	11.35%	13.68%	2.27%	**8.07%**	**0.44**
9			Fair by Rey		5.88%	18.75%	22.21%	33.19%	**13.34%**	**0.78**
10			Huber		6.18%	17.79%	21.55%	19.74%	**11.59%**	**0.60**
11			Logistic Regression		6.05%	18.76%	22.76%	21.88%	**12.10%**	**0.63**
12			Hinch and Talwar		7.85%	11.78%	14.28%	2.31%	**8.22%**	**0.45**
13			Holland and Welsch		7.84%	11.43%	13.82%	2.28%	**8.10%**	**0.44**
14		Regularisation	Lasso		6.70%	2.70%	6.01%	80.53%	**15.22%**	**1.74**
15			Ridge Regression		7.58%	8.93%	15.04%	87.12%	**18.30%**	**1.53**
16			Elastic Nets		6.70%	2.70%	6.01%	80.53%	**15.22%**	**1.74**
17		Stepwise	Interaction	Bounded Steps	8.45%	49.53%	54.40%	12.23%	**17.67%**	**1.21**
18				Unbounded Steps	6.06%	26.78%	35.55%	10.62%	**11.84%**	**1.05**
19			Purequadratic	Bounded Steps	7.58%	2.79%	6.89%	86.49%	**16.49%**	**1.72**
20				Unbounded Steps	5.85%	20.00%	29.53%	11.72%	**10.37%**	**0.93**
21			Quadratic	Bounded Steps	4.34%	3.53%	3.27%	17.75%	**5.78%**	**0.85**
22				Unbounded Steps	7.47%	28.49%	32.13%	11.74%	**12.01%**	**0.97**
23			Polynomial	Bounded Steps	2.26%	2.31%	2.70%	2.34%	**2.12%**	**0.19**
24				Unbounded Steps						

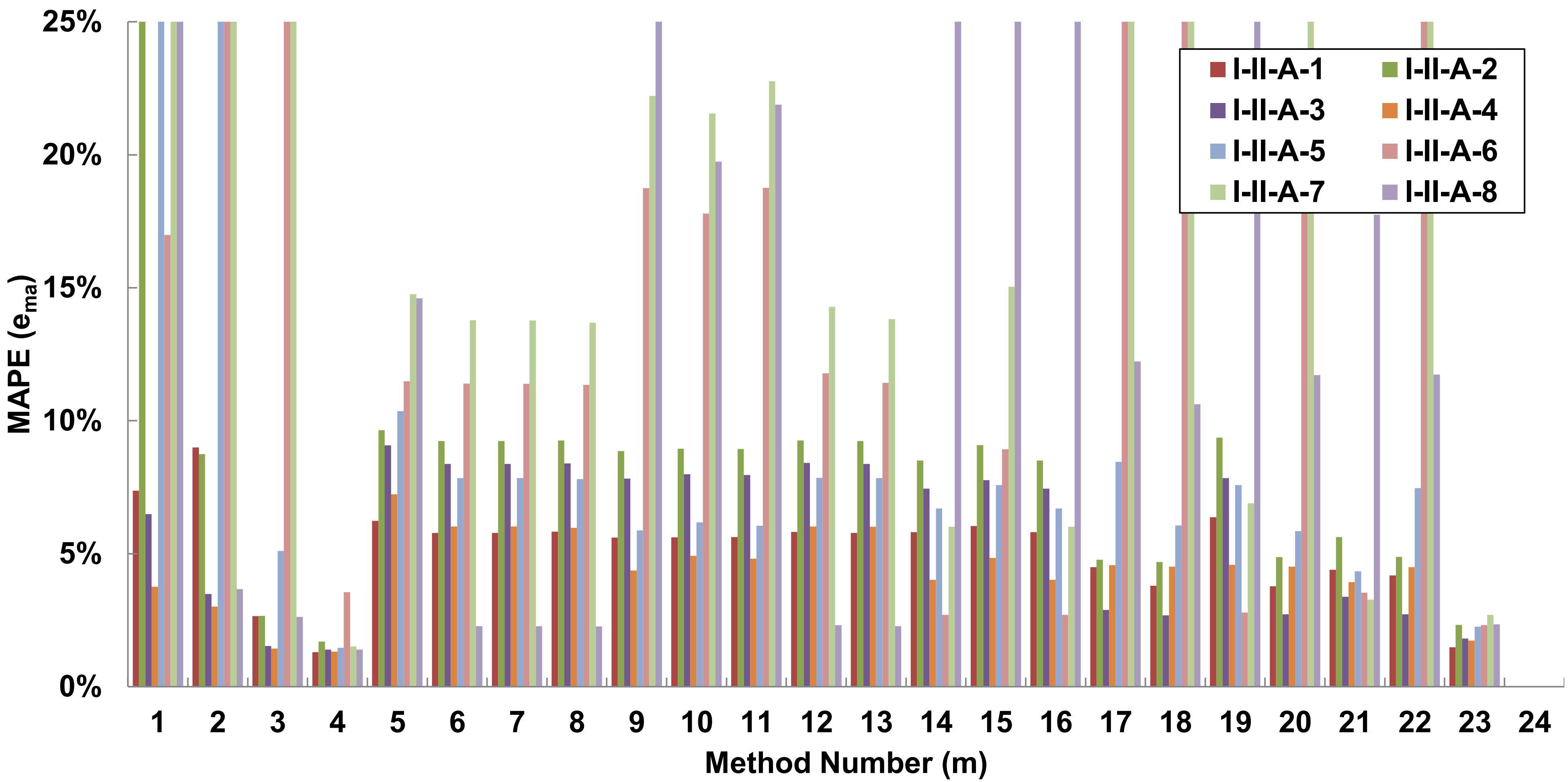


Figure H.38: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}, \log \mu$

Table H.39: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, c

	Method				e_{ma}			
					Test Set $T_o, o=1$	Test Set $T_o, o=1$	Test Set $T_u, u=2$	Test Set $T_u, u=3$
Number (m)	Class I	Class II	Class III	Class IV	*I-II-A-1*	*I-II-A-2*	*I-II-A-3*	*I-II-A-4*
1	Decision Tree	Boosting			6.78%	37.50%	6.01%	4.22%
2	Decision Tree	Bootstrap Aggregating			8.89%	7.90%	2.77%	3.56%
3	Neural Network	Feedforward			4.45%	3.24%	1.88%	1.77%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		83.58%	80.82%	79.50%	23.57%
7	Multiple Linear Regression	Robust	Andrews		83.58%	80.82%	79.50%	23.56%
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		82.80%	80.09%	79.03%	21.34%
9	Multiple Linear Regression	Robust	Fair by Rey		76.29%	73.74%	72.66%	33.88%
10	Multiple Linear Regression	Robust	Huber		78.45%	76.55%	75.73%	25.93%
11	Multiple Linear Regression	Robust	Logistic Regression		78.22%	76.34%	75.54%	26.56%
12	Multiple Linear Regression	Robust	Hinch and Talwar		83.40%	81.15%	79.77%	29.00%
13	Multiple Linear Regression	Robust	Holland and Welsch		83.48%	80.77%	79.46%	23.54%
14	Multiple Linear Regression	Regularisation	Lasso		82.49%	81.60%	80.03%	48.04%
15	Multiple Linear Regression	Regularisation	Ridge Regression		5.91%	9.01%	7.61%	4.57%
16	Multiple Linear Regression	Regularisation	Elastic Nets		82.49%	81.60%	80.03%	48.04%
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	4.35%	5.60%	3.37%	3.93%
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps	48.54%	23.79%	11.54%	16.52%
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	6.37%	9.36%	7.84%	4.58%
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps	26.42%	16.92%	13.17%	23.61%
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	4.40%	5.63%	3.38%	3.93%
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	29.57%	18.36%	13.58%	26.16%
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	1.48%	2.32%	1.81%	1.73%
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps				

Table H.39: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, c (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=3$ *I-II-A-5*	Test Set $T_o, o=4$ *I-II-A-6*	Test Set $T_o, o=5$ *I-II-A-7*	Test Set $T_u, u=1$ *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			44.74%	20.17%	94.97%	71.64%	**35.75%**	**0.94**
2		Bootstrap Aggregating			51.47%	59.54%	35.32%	3.84%	**21.66%**	**1.08**
3	Neural Network	Feedforward			2.62%	91.89%	81.01%	3.60%	**23.81%**	**1.63**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		59.54%	84.22%	76.63%	3.68%	**61.44%**	**0.50**
7			Andrews		59.55%	84.22%	76.63%	3.65%	**61.44%**	**0.50**
8			Cauchy M-estimators by Moore		59.27%	83.71%	76.38%	6.59%	**61.15%**	**0.50**
9			Fair by Rey		99.00%	82.06%	77.25%	51.99%	**70.86%**	**0.28**
10			Huber		75.78%	82.10%	76.42%	42.26%	**66.65%**	**0.31**
11			Logistic Regression		77.30%	82.22%	76.42%	36.79%	**66.17%**	**0.33**
12			Hinch and Talwar		58.39%	84.36%	76.74%	3.58%	**62.05%**	**0.49**
13			Holland and Welsch		59.50%	84.18%	76.62%	3.94%	**61.44%**	**0.50**
14		Regularisation	Lasso		86.16%	30.44%	123.48%	1012.78%	**193.13%**	**1.72**
15			Ridge Regression		7.43%	4.75%	8.36%	86.05%	**16.71%**	**1.68**
16			Elastic Nets		86.16%	30.44%	123.48%	1012.78%	**193.13%**	**1.72**
17		Stepwise	Interaction	Bounded Steps	4.34%	3.53%	3.27%	17.50%	**5.74%**	**0.84**
18				Unbounded Steps	38.92%	425.60%	396.61%	64.70%	**128.28%**	**1.37**
19			Purequadratic	Bounded Steps	7.58%	2.79%	6.89%	86.49%	**16.49%**	**1.72**
20				Unbounded Steps	28.67%	90.58%	84.60%	82.02%	**45.75%**	**0.73**
21			Quadratic	Bounded Steps	4.34%	3.53%	3.27%	17.75%	**5.78%**	**0.85**
22				Unbounded Steps	45.72%	60.32%	56.80%	15.45%	**33.25%**	**0.56**
23			Polynomial	Bounded Steps	2.26%	2.31%	2.70%	2.34%	**2.12%**	**0.19**
24				Unbounded Steps						

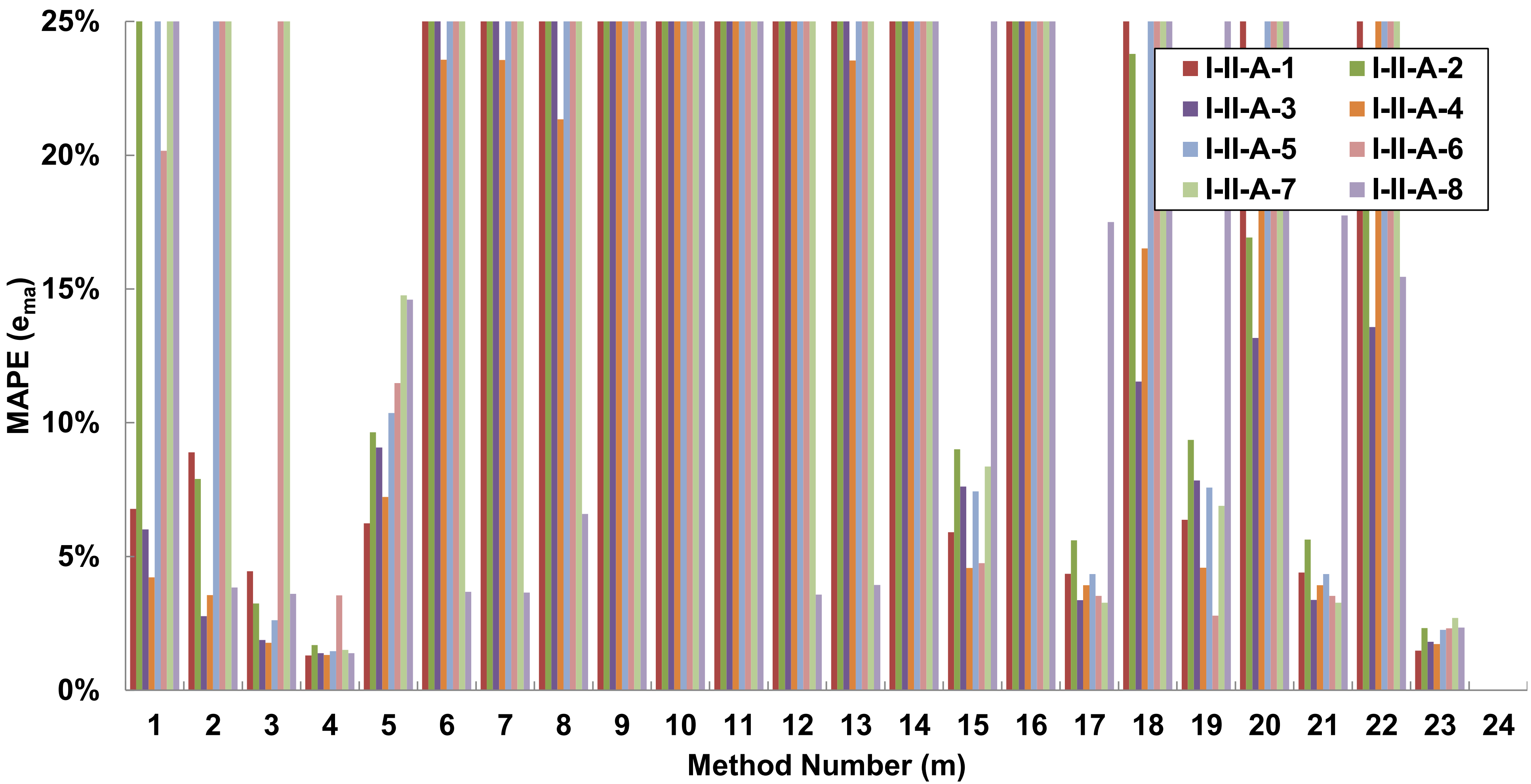


Figure H.39: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}},c$

Table H.40: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, c^{-1}

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ I-II-A-1	Test Set $T_o, o=1$ I-II-A-2	Test Set $T_u, u=2$ I-II-A-3	Test Set $T_u, u=3$ I-II-A-4
1	Decision Tree	Boosting			6.90%	39.81%	6.02%	4.26%
2		Bootstrap Aggregating			9.39%	8.17%	2.92%	3.42%
3	Neural Network	Feedforward			33.45%	3.54%	1.62%	1.77%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		82.68%	80.80%	79.50%	23.57%
7			Andrews		82.68%	80.80%	79.50%	23.56%
8			Cauchy M-estimators by Moore		82.65%	80.09%	79.03%	21.34%
9			Fair by Rey		75.96%	73.73%	72.66%	33.88%
10			Huber		78.14%	76.54%	75.73%	25.93%
11			Logistic Regression		77.93%	76.34%	75.54%	26.56%
12			Hinch and Talwar		81.75%	81.12%	79.77%	29.00%
13			Holland and Welsch		82.58%	80.75%	79.46%	23.54%
14		Regularisation	Lasso		79.44%	81.54%	80.03%	48.04%
15			Ridge Regression		5.91%	9.01%	7.61%	4.57%
16			Elastic Nets		79.44%	81.54%	80.03%	48.04%
17		Stepwise	Interaction	Bounded Steps	4.35%	5.60%	3.37%	3.93%
18				Unbounded Steps				
19			Purequadratic	Bounded Steps	6.37%	9.36%	7.84%	4.58%
20				Unbounded Steps				
21			Quadratic	Bounded Steps	4.40%	5.63%	3.38%	3.93%
22				Unbounded Steps	28.32%	18.35%	13.58%	26.16%
23			Polynomial	Bounded Steps	1.48%	2.32%	1.81%	1.73%
24				Unbounded Steps				

Table H.40: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}, c^{-1}$ (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_ma: Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			43.53%	29.65%	110.70%	72.26%	**39.14%**	**0.95**
2	Decision Tree	Bootstrap Aggregating			47.84%	65.28%	44.49%	3.67%	**23.15%**	**1.09**
3	Neural Network	Feedforward			5.44%	58.45%	50.89%	2.92%	**19.76%**	**1.22**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		59.54%	84.22%	76.63%	3.68%	**61.33%**	**0.50**
7	Multiple Linear Regression	Robust	Andrews		59.55%	84.22%	76.63%	3.65%	**61.32%**	**0.50**
8	Multiple Linear Regression	Robust	Cauchy M-estimators by Moore		59.27%	83.71%	76.38%	6.59%	**61.13%**	**0.50**
9	Multiple Linear Regression	Robust	Fair by Rey		99.00%	82.06%	77.25%	51.99%	**70.82%**	**0.28**
10	Multiple Linear Regression	Robust	Huber		75.78%	82.10%	76.42%	42.26%	**66.61%**	**0.31**
11	Multiple Linear Regression	Robust	Logistic Regression		77.30%	82.22%	76.42%	36.79%	**66.14%**	**0.33**
12	Multiple Linear Regression	Robust	Hinch and Talwar		58.39%	84.36%	76.74%	3.58%	**61.84%**	**0.49**
13	Multiple Linear Regression	Robust	Holland and Welsch		59.50%	84.18%	76.62%	3.94%	**61.32%**	**0.50**
14	Multiple Linear Regression	Regularisation	Lasso		86.16%	30.44%	123.48%	1012.78%	**192.74%**	**1.73**
15	Multiple Linear Regression	Regularisation	Ridge Regression		7.43%	4.75%	8.36%	86.05%	**16.71%**	**1.68**
16	Multiple Linear Regression	Regularisation	Elastic Nets		86.16%	30.44%	123.48%	1012.78%	**192.74%**	**1.73**
17	Multiple Linear Regression	Stepwise	Interaction	Bounded Steps	4.34%	3.53%	3.27%	17.50%	**5.74%**	**0.84**
18	Multiple Linear Regression	Stepwise	Interaction	Unbounded Steps						
19	Multiple Linear Regression	Stepwise	Purequadratic	Bounded Steps	7.58%	2.79%	6.89%	86.49%	**16.49%**	**1.72**
20	Multiple Linear Regression	Stepwise	Purequadratic	Unbounded Steps						
21	Multiple Linear Regression	Stepwise	Quadratic	Bounded Steps	4.34%	3.53%	3.27%	17.75%	**5.78%**	**0.85**
22	Multiple Linear Regression	Stepwise	Quadratic	Unbounded Steps	45.72%	60.32%	56.80%	15.45%	**33.09%**	**0.56**
23	Multiple Linear Regression	Stepwise	Polynomial	Bounded Steps	2.26%	2.31%	2.70%	2.34%	**2.12%**	**0.19**
24	Multiple Linear Regression	Stepwise	Polynomial	Unbounded Steps						

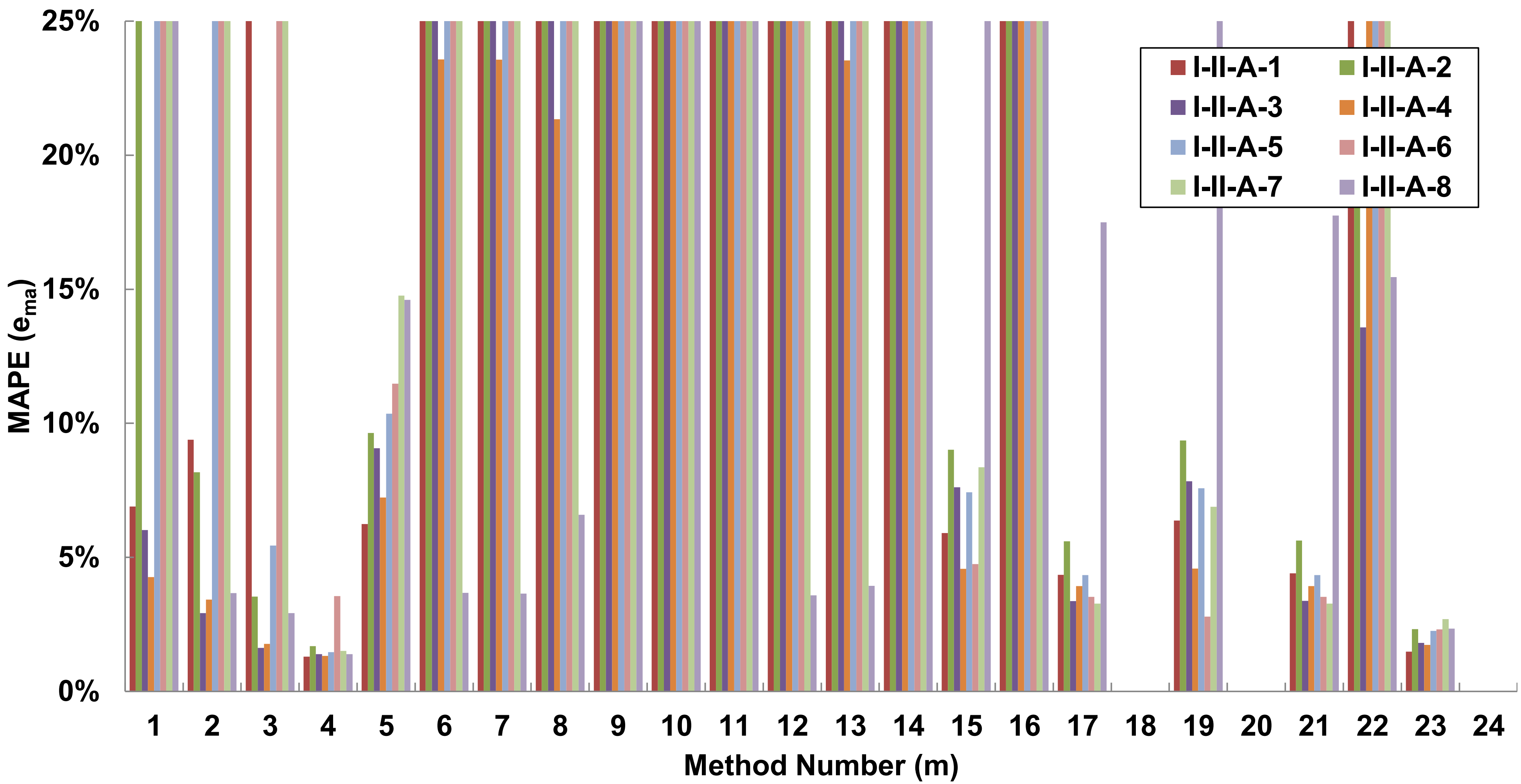


Figure H.40: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}, c^{-1}$

Table H.41: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, $\log c$

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set $T_o, o=1$ I-II-A-1	Test Set $T_o, o=1$ I-II-A-2	Test Set $T_u, u=2$ I-II-A-3	Test Set $T_u, u=3$ I-II-A-4
1	Decision Tree	Boosting			6.37%	41.02%	6.03%	4.20%
2	Decision Tree	Bootstrap Aggregating			8.05%	8.39%	2.80%	3.50%
3	Neural Network	Feedforward			4.00%	3.28%	1.80%	2.44%
4	Li and Meerkov (2009)				1.30%	1.69%	1.39%	1.32%
5	$1/\mu_{\max}$				6.24%	9.64%	9.07%	7.23%
6	Multiple Linear Regression	Robust	Tukey's Bisquare		4.80%	9.21%	8.38%	6.04%
7			Andrews		4.80%	9.21%	8.38%	6.04%
8			Cauchy M-estimators by Moore		4.83%	9.23%	8.38%	5.99%
9			Fair by Rey		4.77%	8.84%	7.90%	4.44%
10			Huber		4.61%	8.94%	8.05%	4.99%
11			Logistic Regression		4.61%	8.93%	8.04%	4.94%
12			Hinch and Talwar		4.80%	9.21%	8.38%	6.04%
13			Holland and Welsch		4.80%	9.22%	8.38%	6.04%
14		Regularisation	Lasso		5.33%	8.54%	7.50%	3.93%
15			Ridge Regression		5.17%	8.99%	7.61%	4.57%
16			Elastic Nets		5.33%	8.54%	7.50%	3.93%
17		Stepwise	Interaction	Bounded Steps	3.65%	5.27%	3.21%	3.67%
18				Unbounded Steps	3.29%	5.75%	2.90%	4.10%
19			Purequadratic	Bounded Steps	5.71%	9.36%	7.85%	4.10%
20				Unbounded Steps	3.31%	5.34%	2.91%	4.01%
21			Quadratic	Bounded Steps	3.67%	5.31%	3.21%	3.70%
22				Unbounded Steps	3.33%	6.00%	2.89%	3.99%
23			Polynomial	Bounded Steps	1.13%	2.31%	1.81%	1.73%
24				Unbounded Steps				

Table H.41: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}$, $\log c$ (*cont.*)

Number (m)	Class I	Class II	Class III	Class IV	e_{ma} Test Set T_o, o=3 *I-II-A-5*	Test Set T_o, o=4 *I-II-A-6*	Test Set T_o, o=5 *I-II-A-7*	Test Set T_u, u=1 *I-II-A-8*	μ_e	c_e
1	Decision Tree	Boosting			43.83%	21.18%	97.12%	71.88%	**36.45%**	**0.93**
2		Bootstrap Aggregating			48.13%	69.15%	50.41%	3.67%	**24.26%**	**1.11**
3	Neural Network	Feedforward			4.76%	40.09%	38.62%	9.79%	**13.10%**	**1.25**
4	Li and Meerkov (2009)				1.46%	3.55%	1.51%	1.39%	**1.70%**	**0.45**
5	$1/\mu_{\max}$				10.36%	11.48%	14.76%	14.60%	**10.42%**	**0.30**
6	Multiple Linear Regression	Robust	Tukey's Bisquare		7.86%	9.74%	11.56%	2.29%	**7.49%**	**0.40**
7			Andrews		7.86%	9.74%	11.56%	2.29%	**7.49%**	**0.40**
8			Cauchy M-estimators by Moore		7.80%	9.75%	11.54%	2.35%	**7.48%**	**0.40**
9			Fair by Rey		4.63%	7.18%	6.69%	32.24%	**9.59%**	**0.97**
10			Huber		5.94%	9.07%	9.65%	17.37%	**8.58%**	**0.47**
11			Logistic Regression		5.85%	8.83%	9.26%	18.08%	**8.57%**	**0.50**
12			Hinch and Talwar		7.85%	9.74%	11.55%	2.29%	**7.48%**	**0.40**
13			Holland and Welsch		7.85%	9.74%	11.56%	2.29%	**7.49%**	**0.40**
14		Regularisation	Lasso		6.25%	3.31%	6.67%	81.26%	**15.35%**	**1.74**
15			Ridge Regression		7.43%	4.75%	8.36%	86.05%	**16.62%**	**1.69**
16			Elastic Nets		6.25%	3.31%	6.67%	81.26%	**15.35%**	**1.74**
17		Stepwise	Interaction	Bounded Steps	3.95%	4.68%	4.48%	9.10%	**4.75%**	**0.40**
18				Unbounded Steps	15.41%	31.93%	31.20%	3.69%	**12.28%**	**1.02**
19			Purequadratic	Bounded Steps	7.31%	3.21%	7.90%	89.27%	**16.84%**	**1.74**
20				Unbounded Steps	15.09%	6.52%	6.90%	5.64%	**6.22%**	**0.62**
21			Quadratic	Bounded Steps	3.95%	4.66%	4.51%	8.77%	**4.72%**	**0.37**
22				Unbounded Steps	15.02%	47.87%	46.10%	5.75%	**16.37%**	**1.18**
23			Polynomial	Bounded Steps	2.26%	2.31%	2.70%	2.34%	**2.07%**	**0.24**
24				Unbounded Steps						

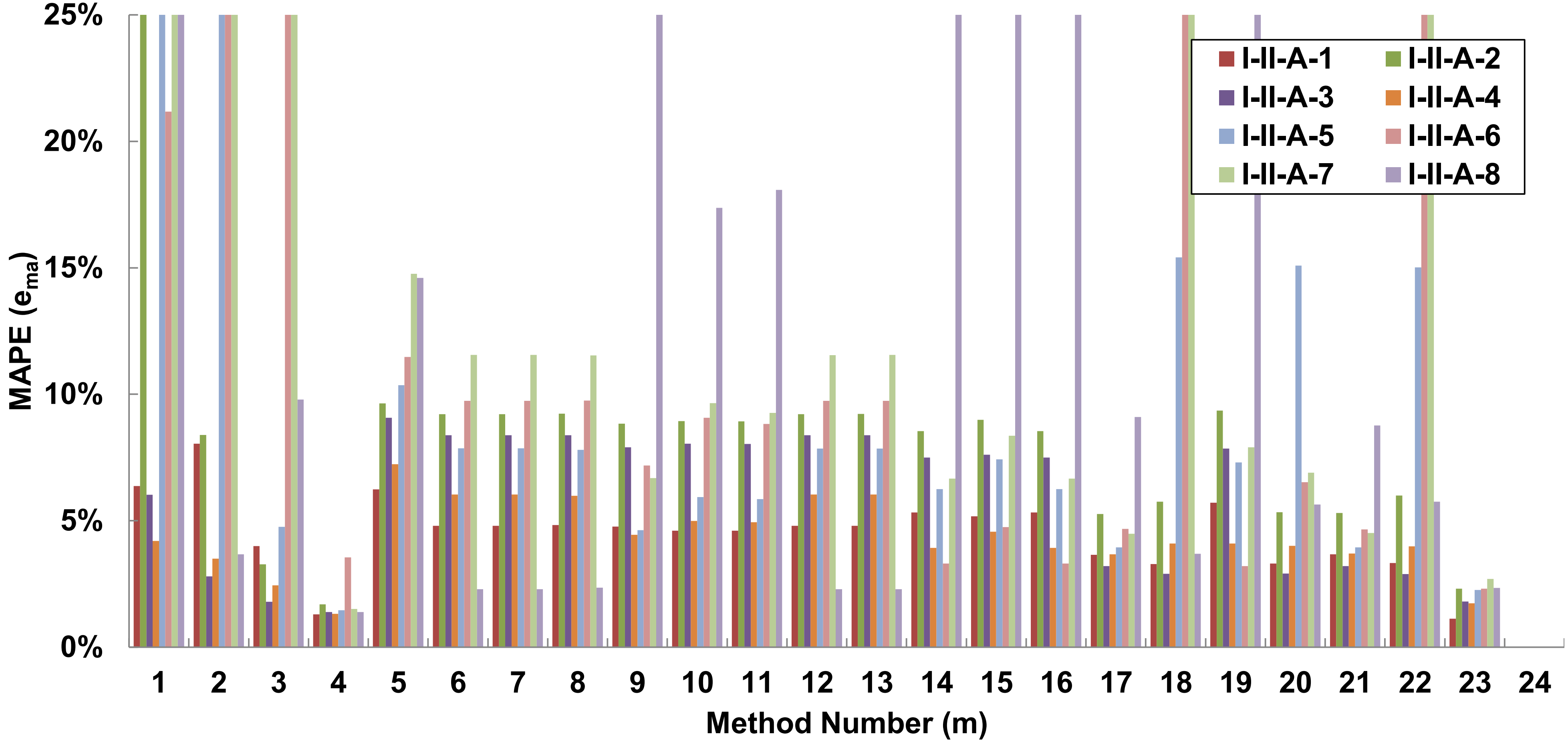


Figure H.41: MAPE of the Individual Test Set of Data Mining Models for Asynchronous Flow Lines with the Supp. Predictor $e^{\mu^{-1}}, \log c$

APPENDIX I – Phase III – Residual Plots for the Optimal Regression Model for the Throughput Rate of Synchronous Flow Lines

Refer to Section 6.4.1

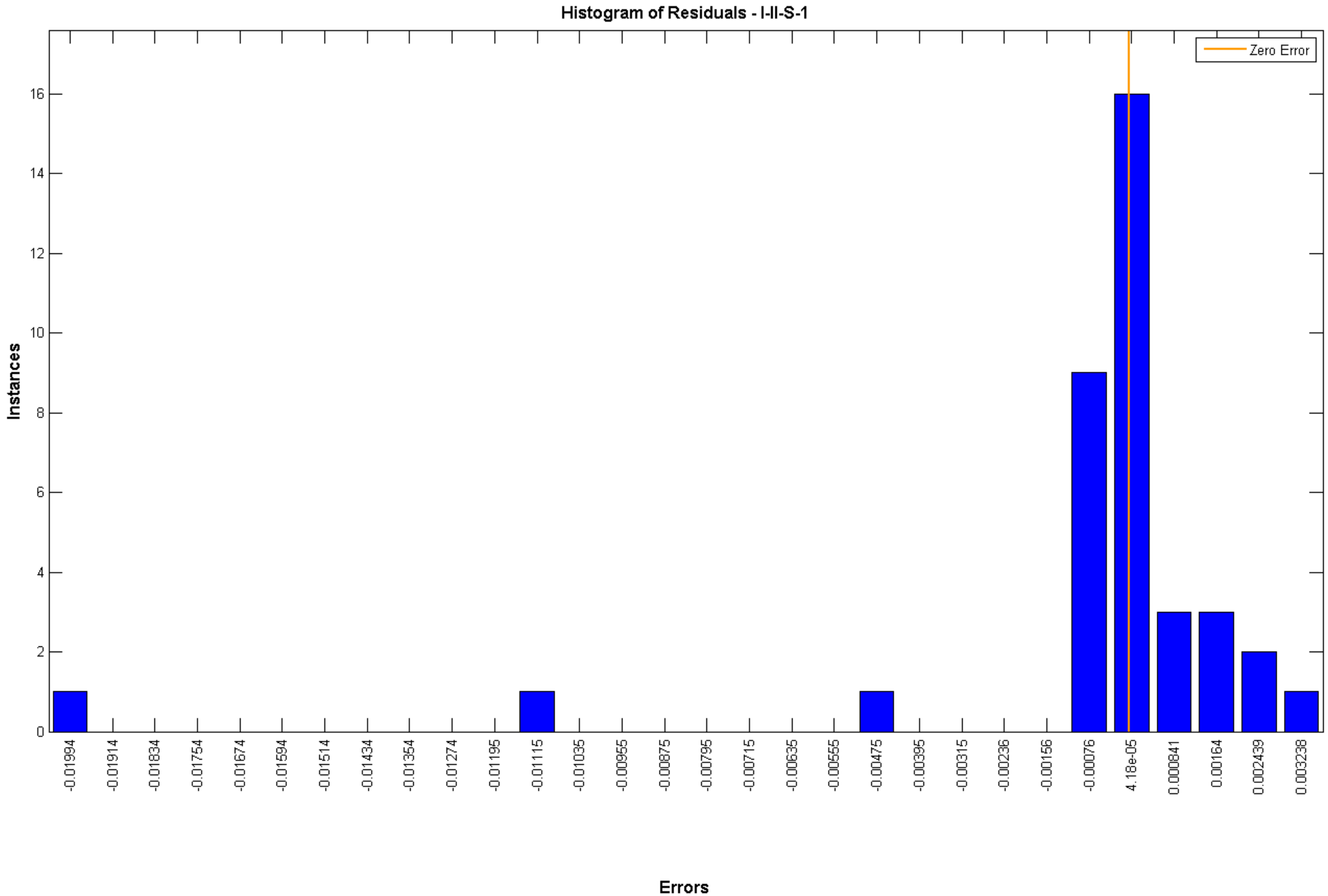


Figure I.1: Histogram of Errors using the Optimal Regression Model for Test Data Set $I-II-S-1$

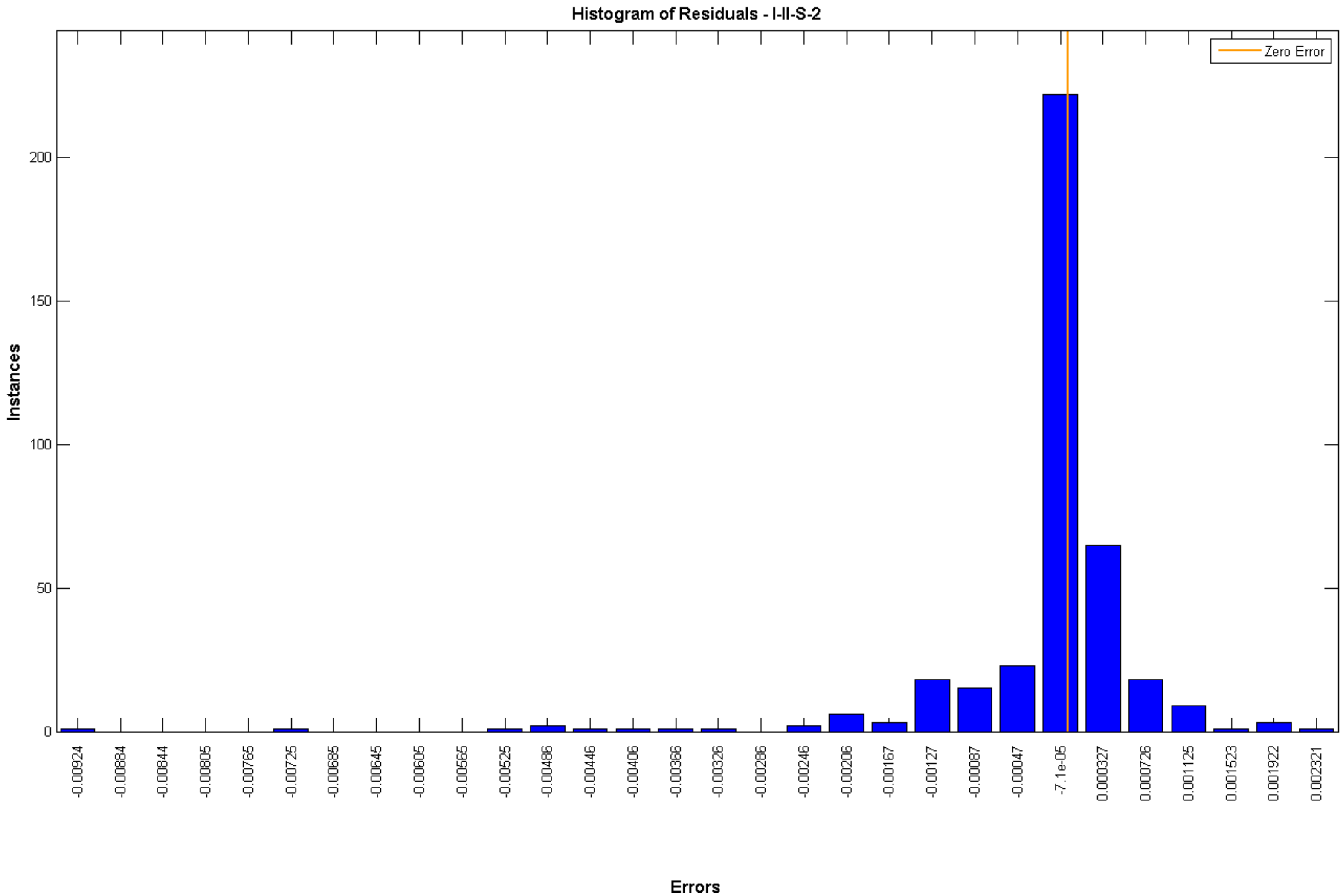


Figure I.2: Histogram of Errors using the Optimal Regression Model for Test Data Set $I-II-S-2$

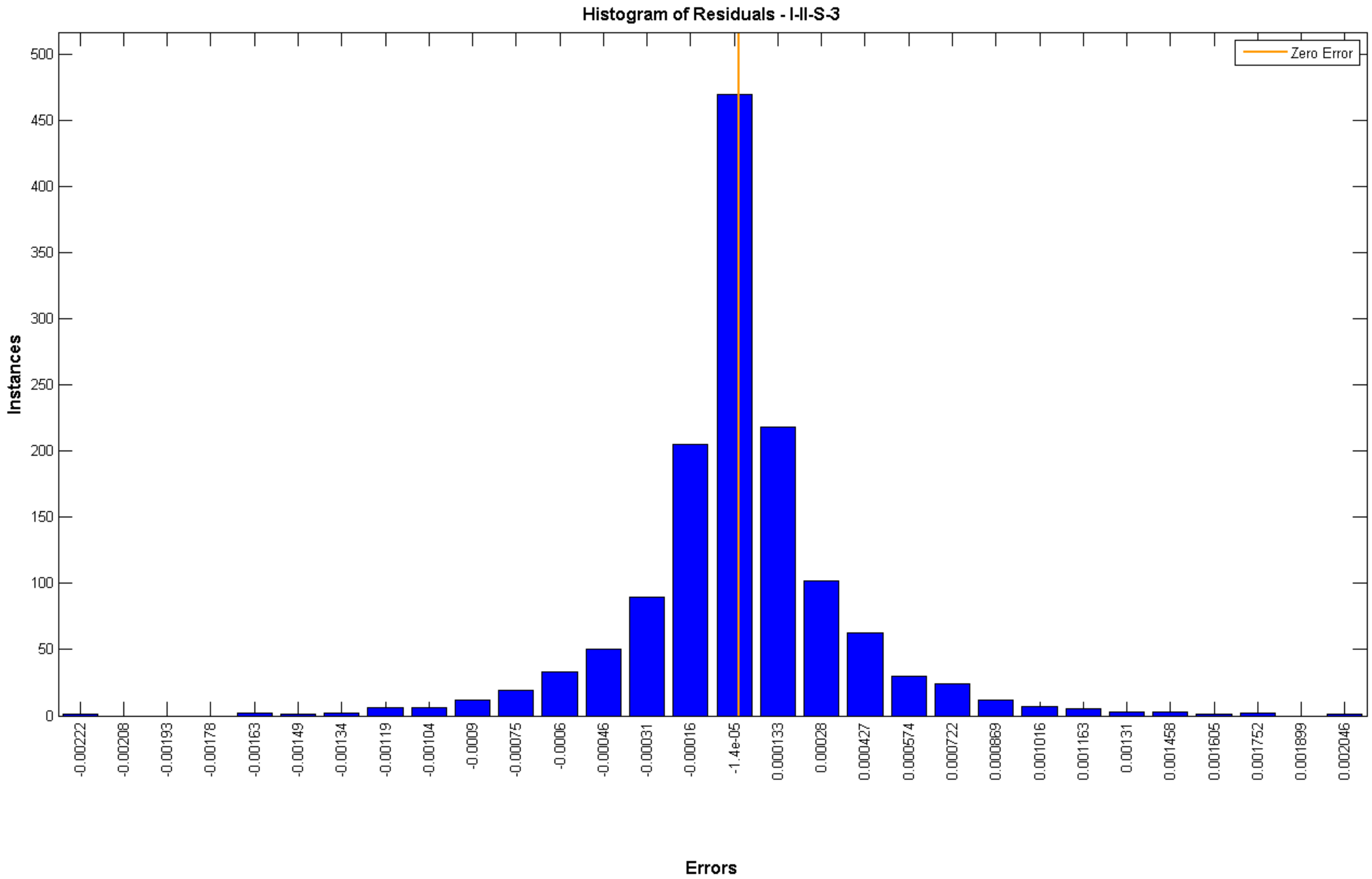


Figure I.3: Histogram of Errors using the Optimal Regression Model for Test Data Set $I-II-S-3$

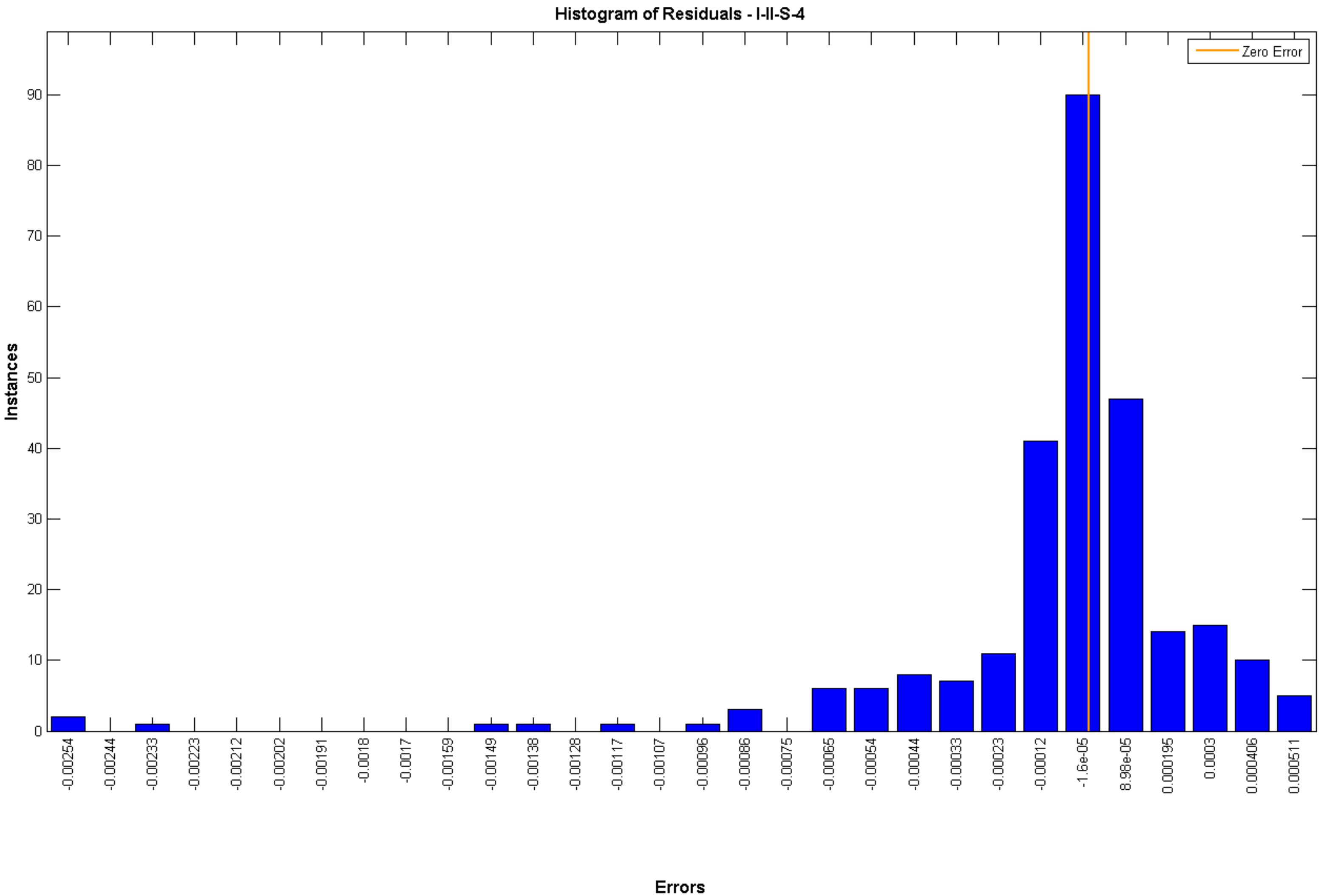


Figure I.4: Histogram of Errors using the Optimal Regression Model for Test Data Set $I-II-S-4$

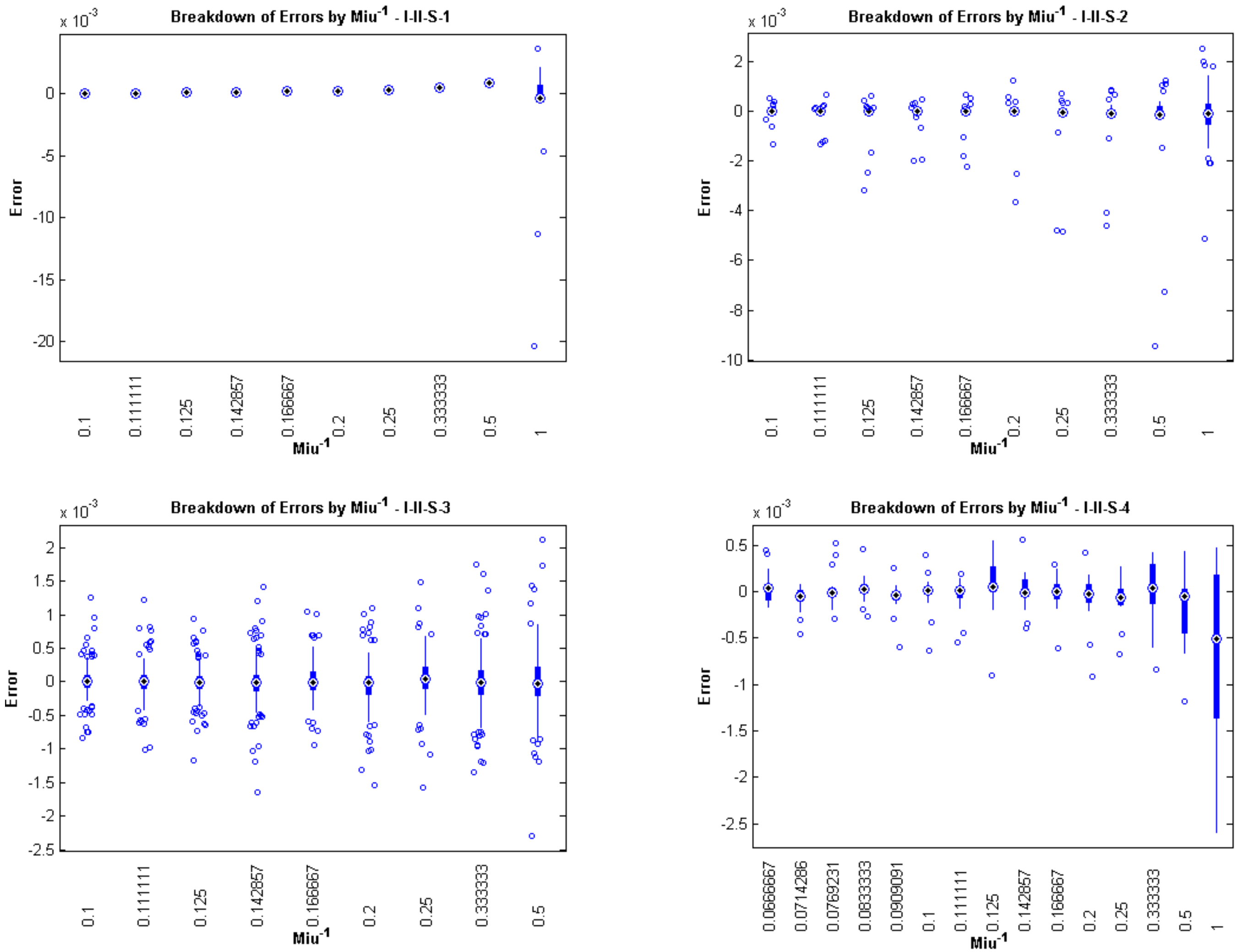


Figure I.5: Breakdown of Errors by μ^{-1} using the Optimal Regression Model for all Data Sets

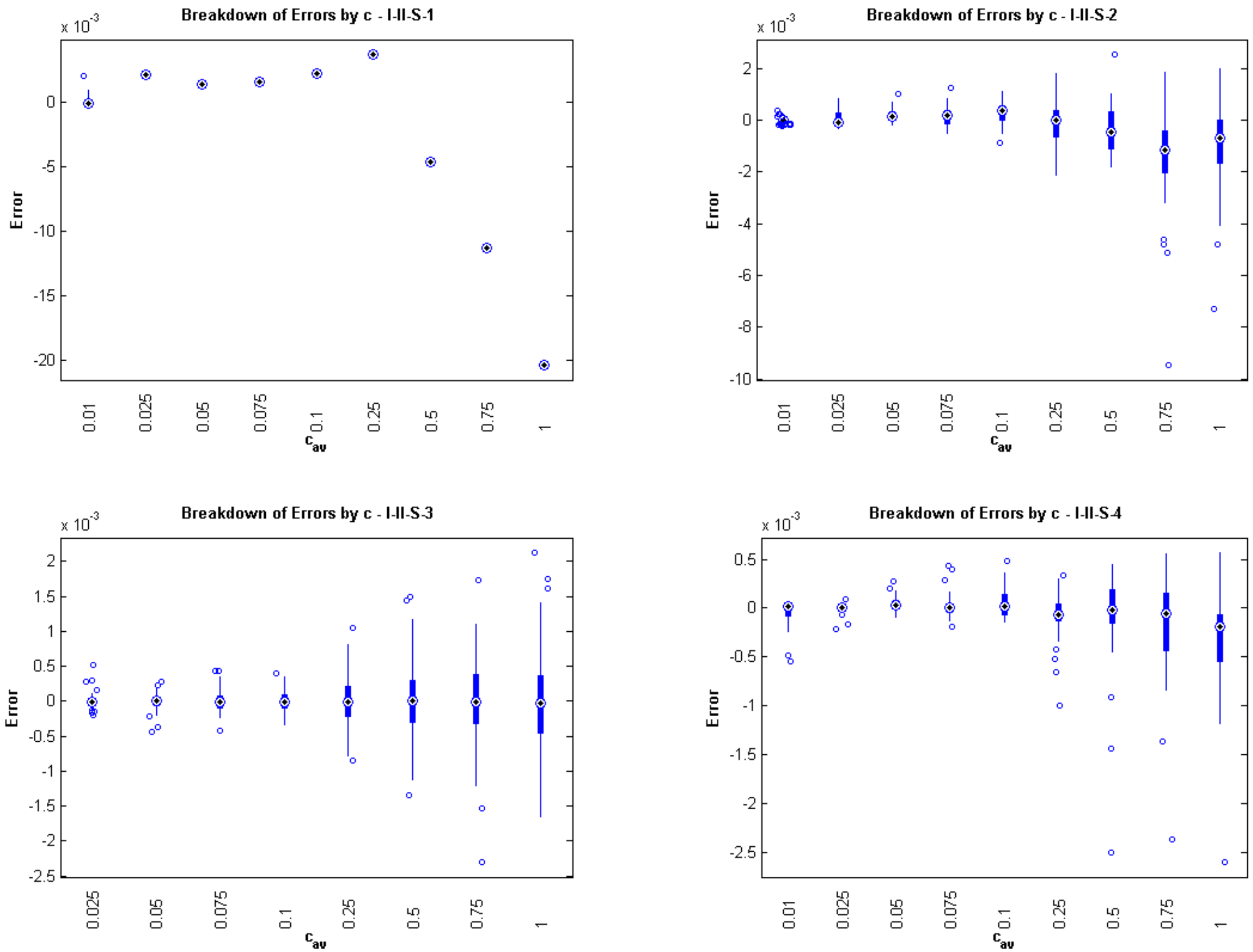


Figure I.6: Breakdown of Errors by c using the Optimal Regression Model for all Data Sets

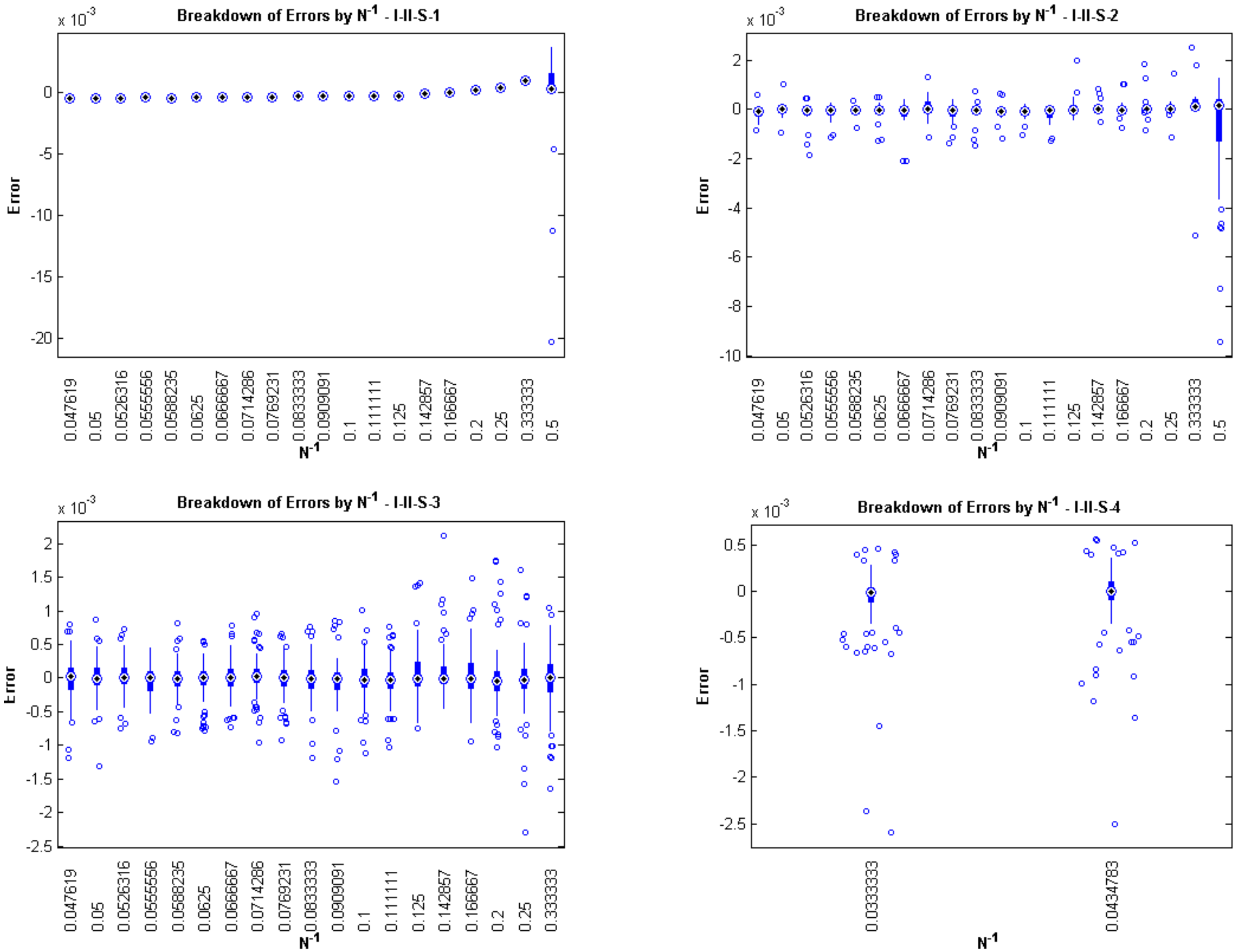


Figure I.7: Breakdown of Errors by N^{-1} using the Optimal Regression Model for all Data Sets

APPENDIX J – Phase III – Residual Plots for the Optimal Regression Model for the Throughput Rate of Asynchronous Flow Lines

Refer to Section 6.4.2.2.1

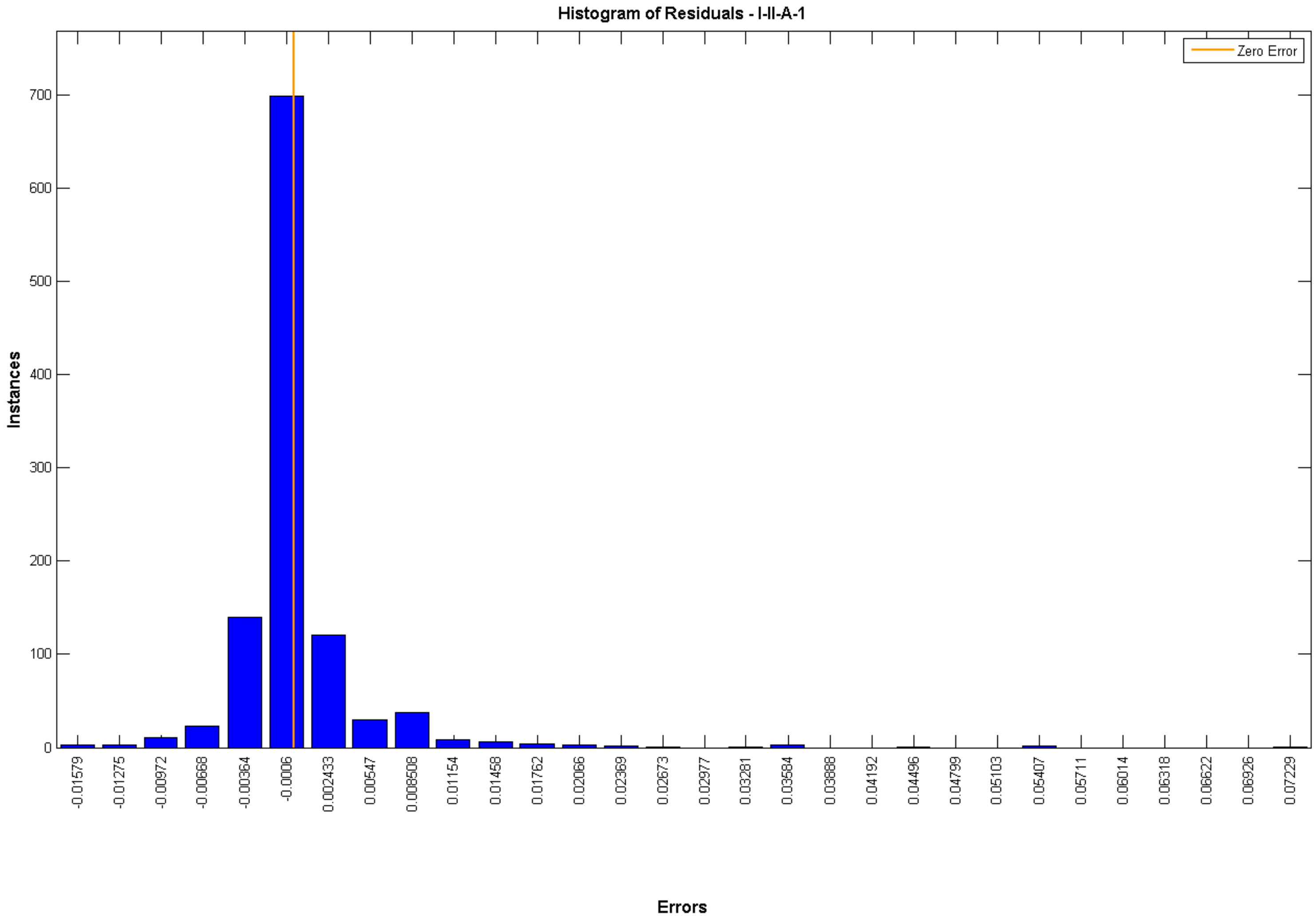


Figure J.1: Histogram of Errors using the Optimal Regression Model for Test Data Set $I-II-A-1$

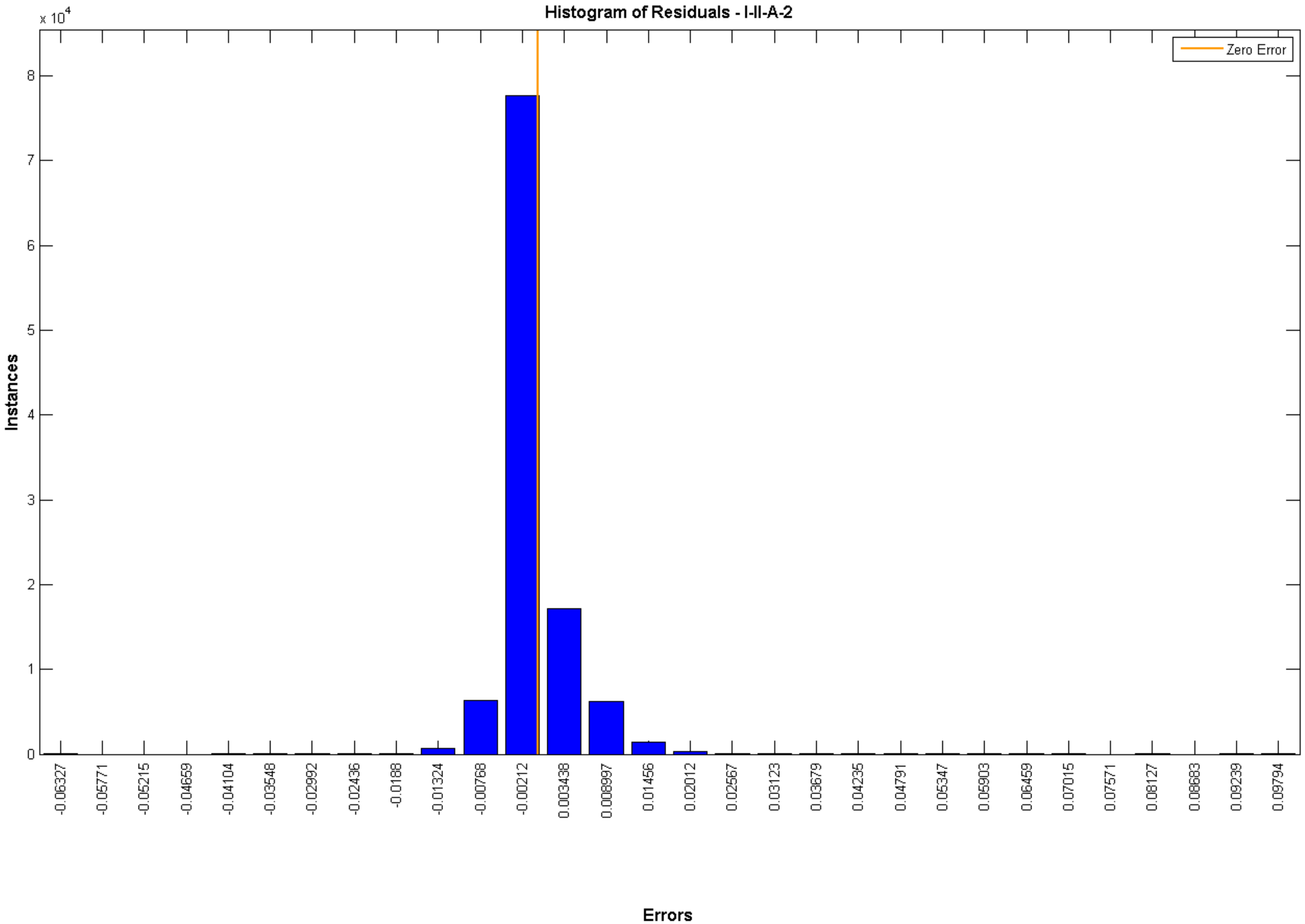


Figure J.2: Histogram of Errors using the Optimal Regression Model for Test Data Set $I-II-A-2$

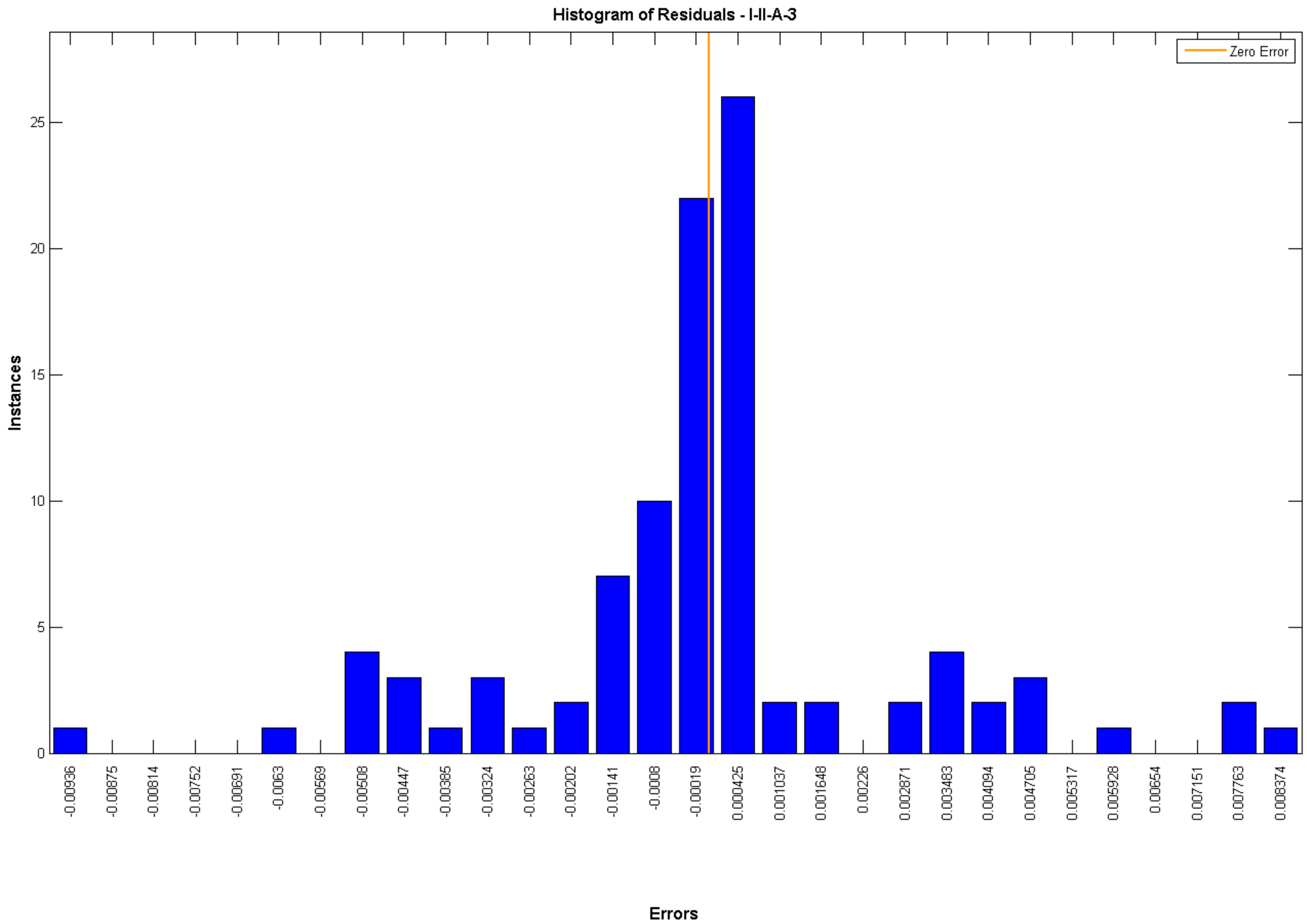


Figure J.3: Histogram of Errors using the Optimal Regression Model for Test Data Set $I-II-A-3$

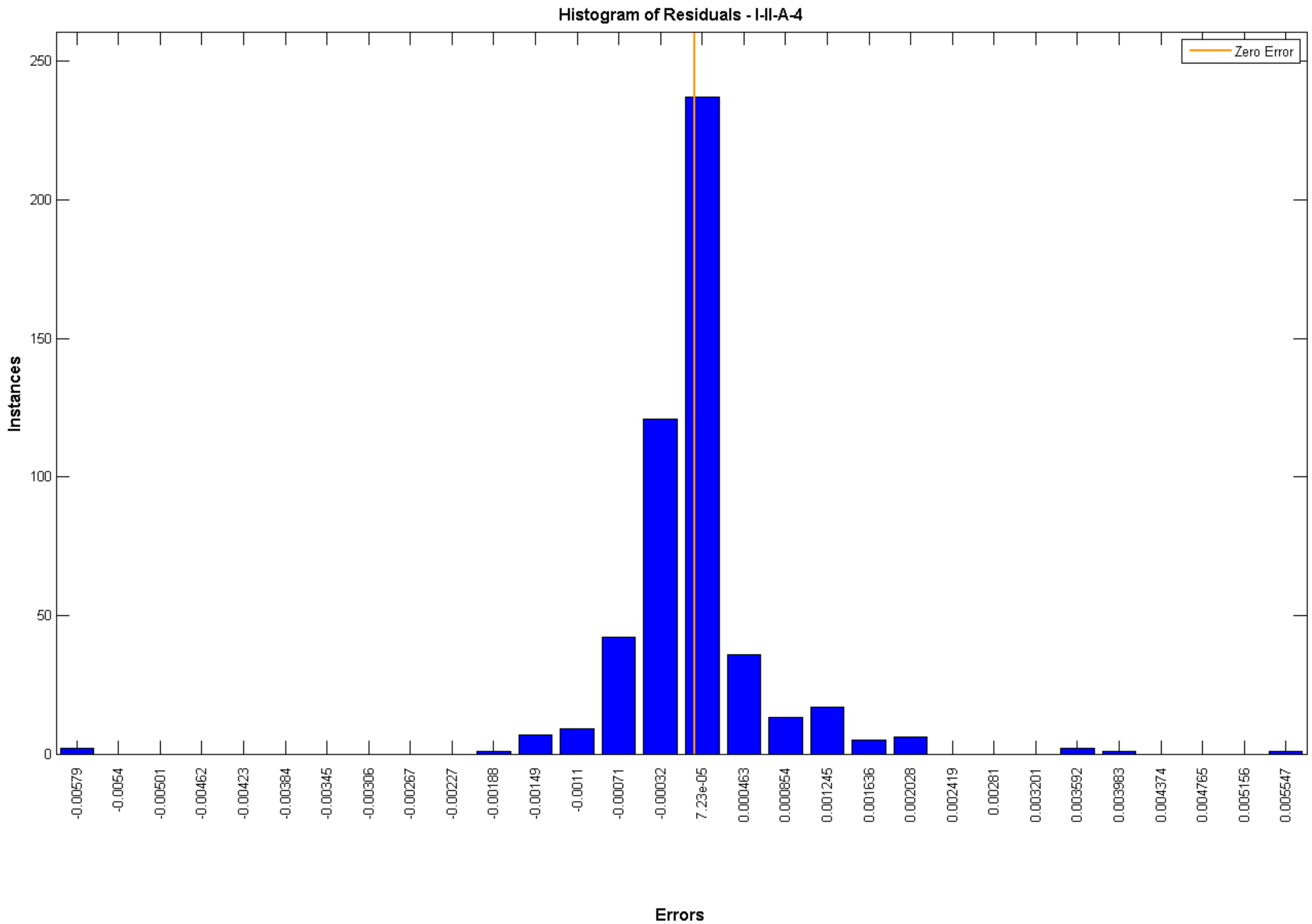


Figure J.4: Histogram of Errors using the Optimal Regression Model for Test Data Set $I-II-A-4$

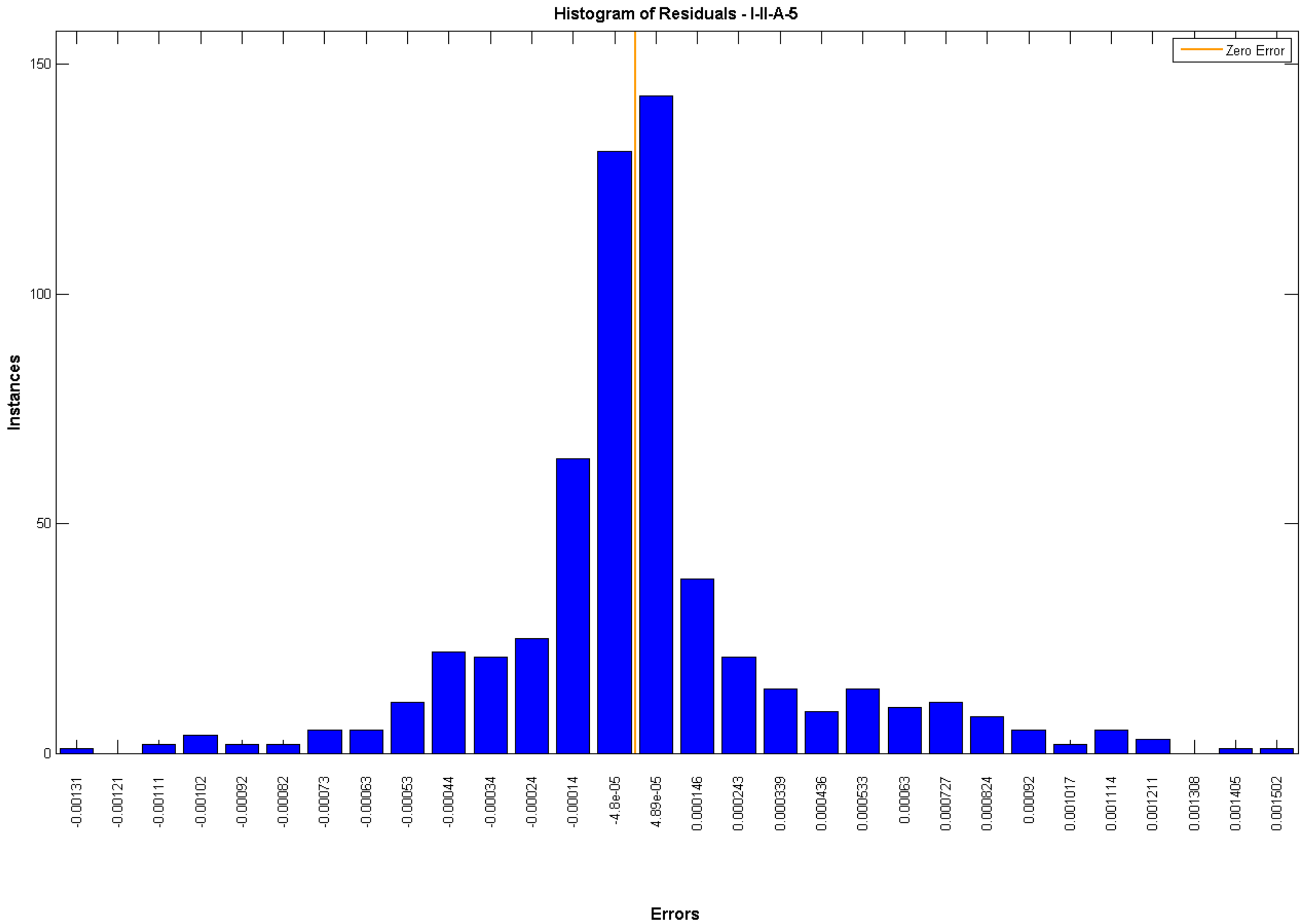


Figure J.5: Histogram of Errors using the Optimal Regression Model for Test Data Set $I-II-A-5$

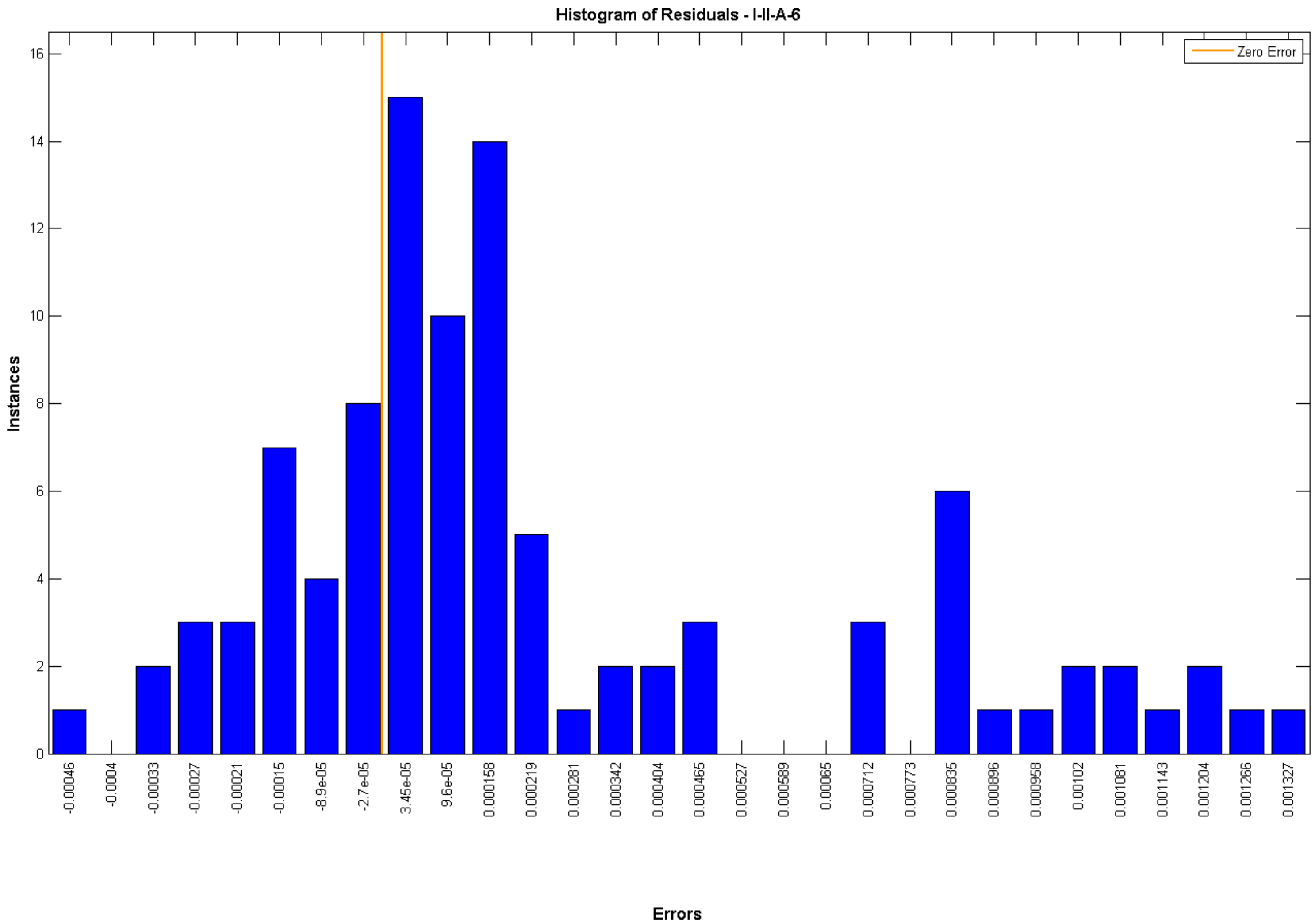


Figure J.6: Histogram of Errors using the Optimal Regression Model for Test Data Set $I - II - A - 6$

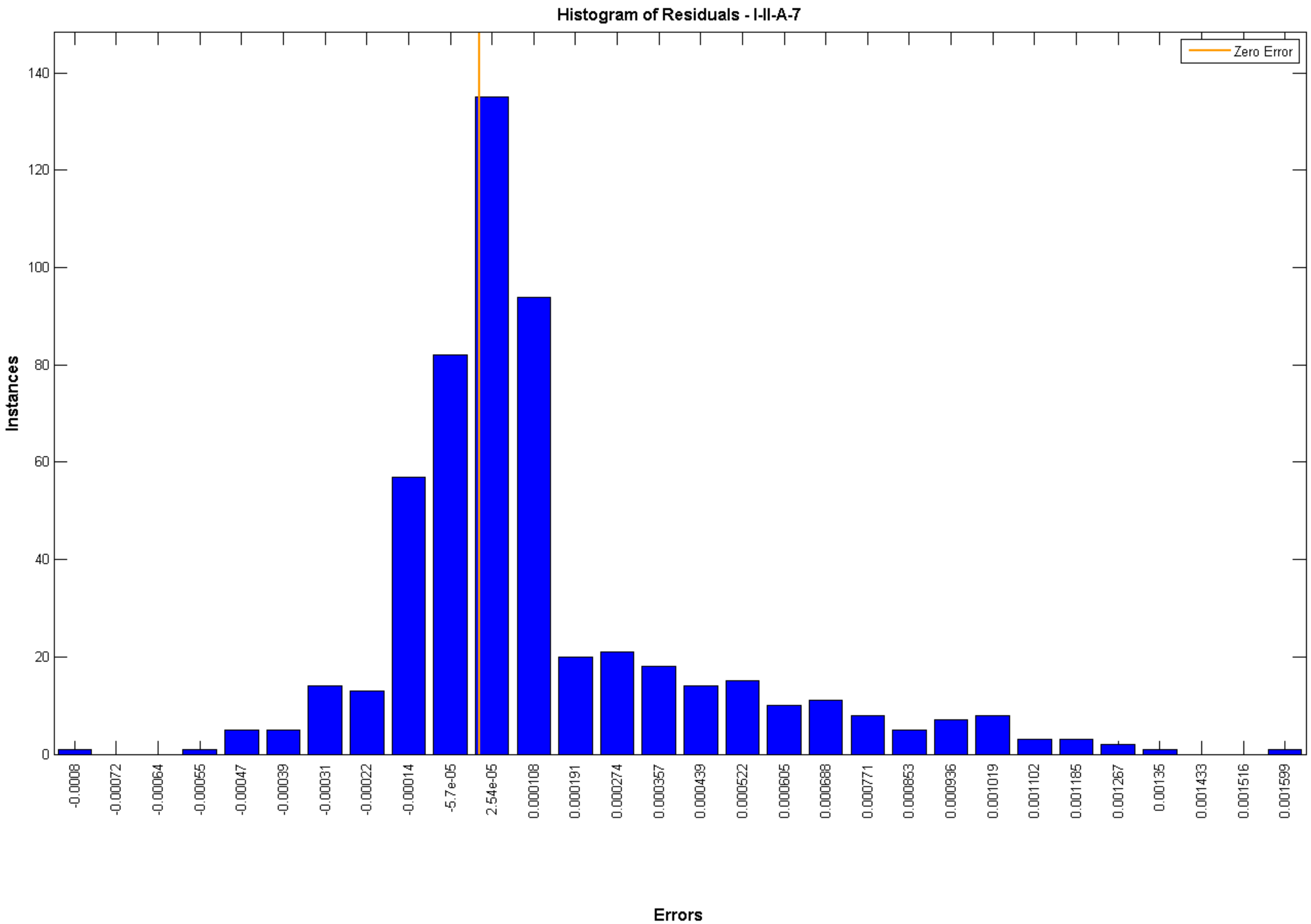


Figure J.7: Histogram of Errors using the Optimal Regression Model for Test Data Set $I-II-A-7$

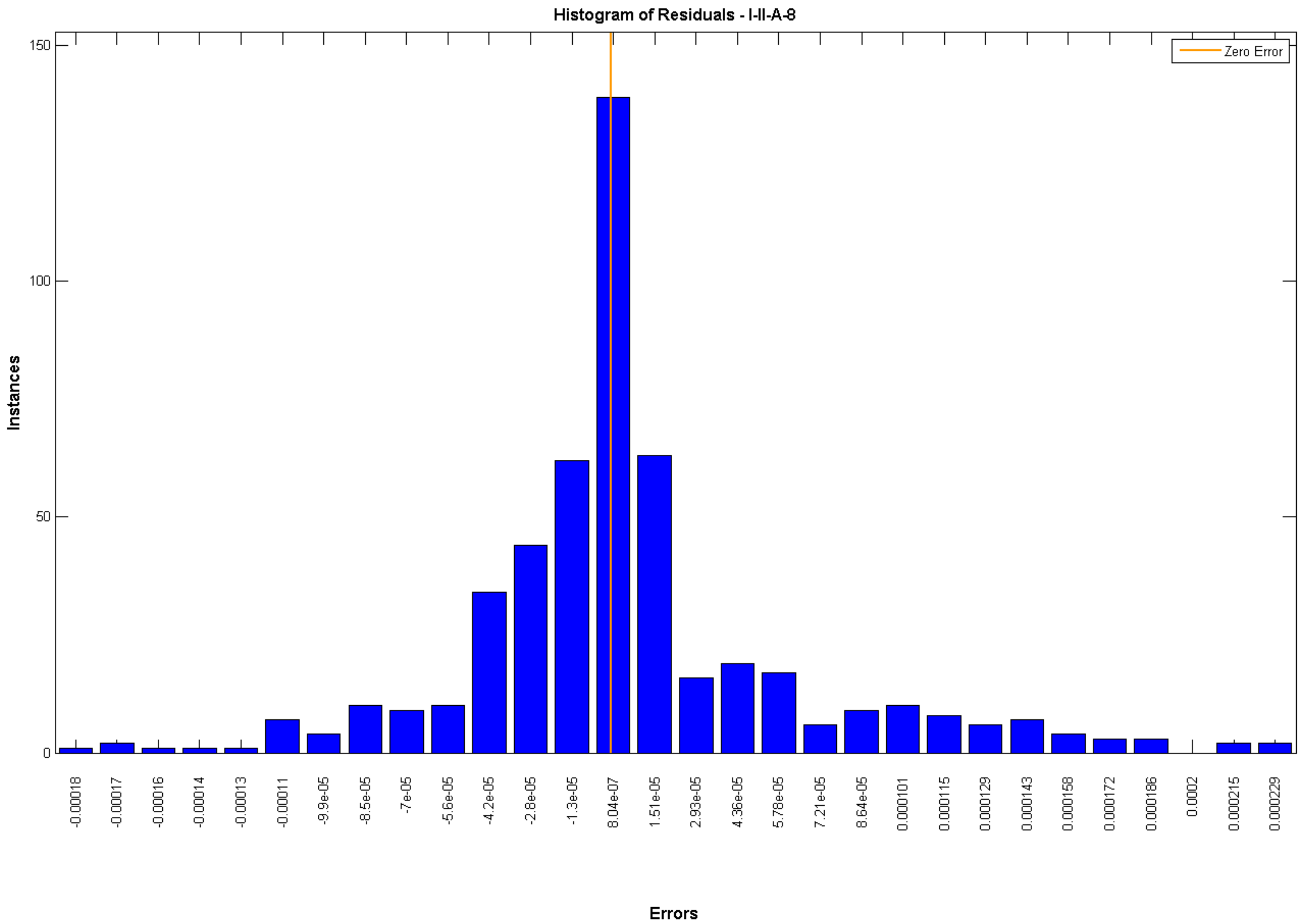


Figure J.8: Histogram of Errors using the Optimal Regression Model for Test Data Set $I-II-A-8$

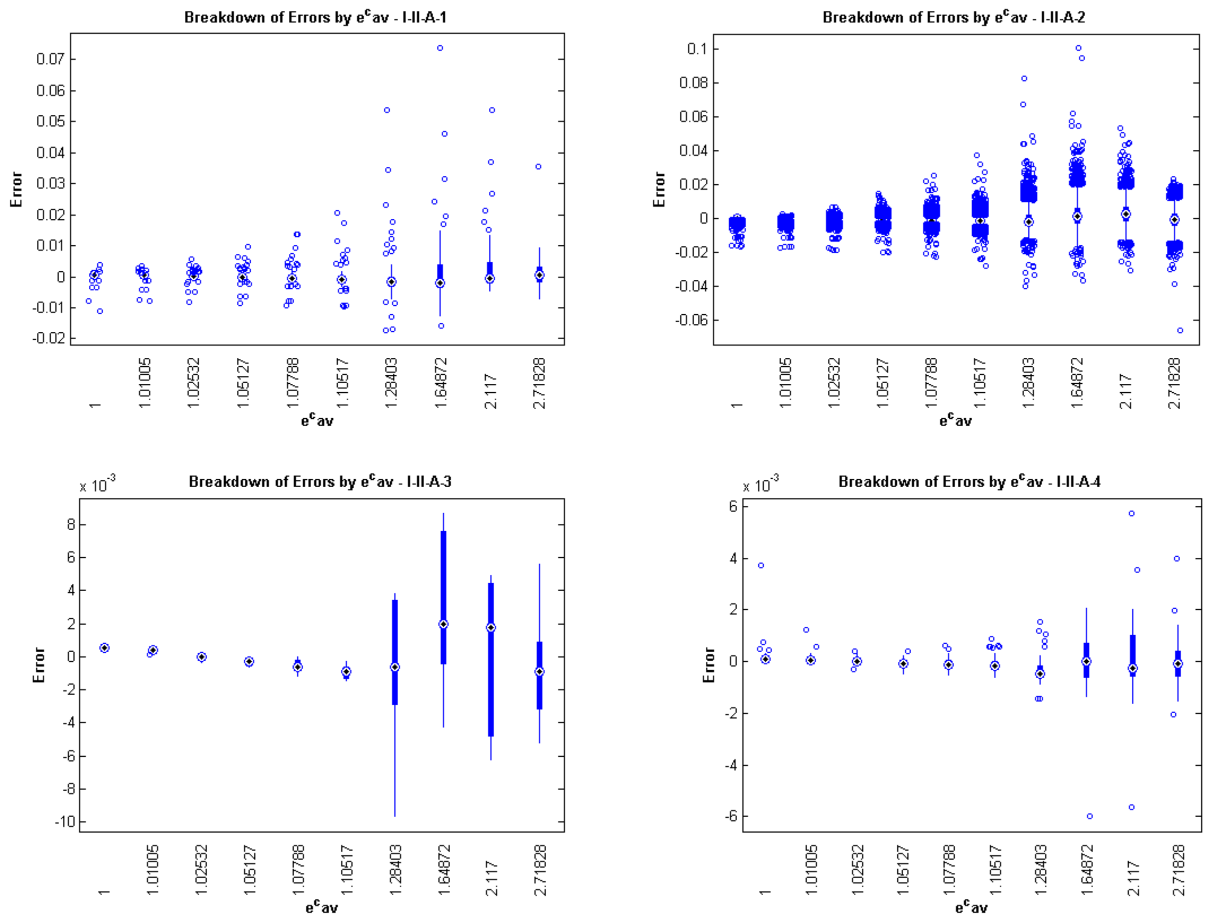


Figure J.9: Breakdown of Errors by $e^{c_{av}}$ using the Optimal Regression Model for all Data Sets

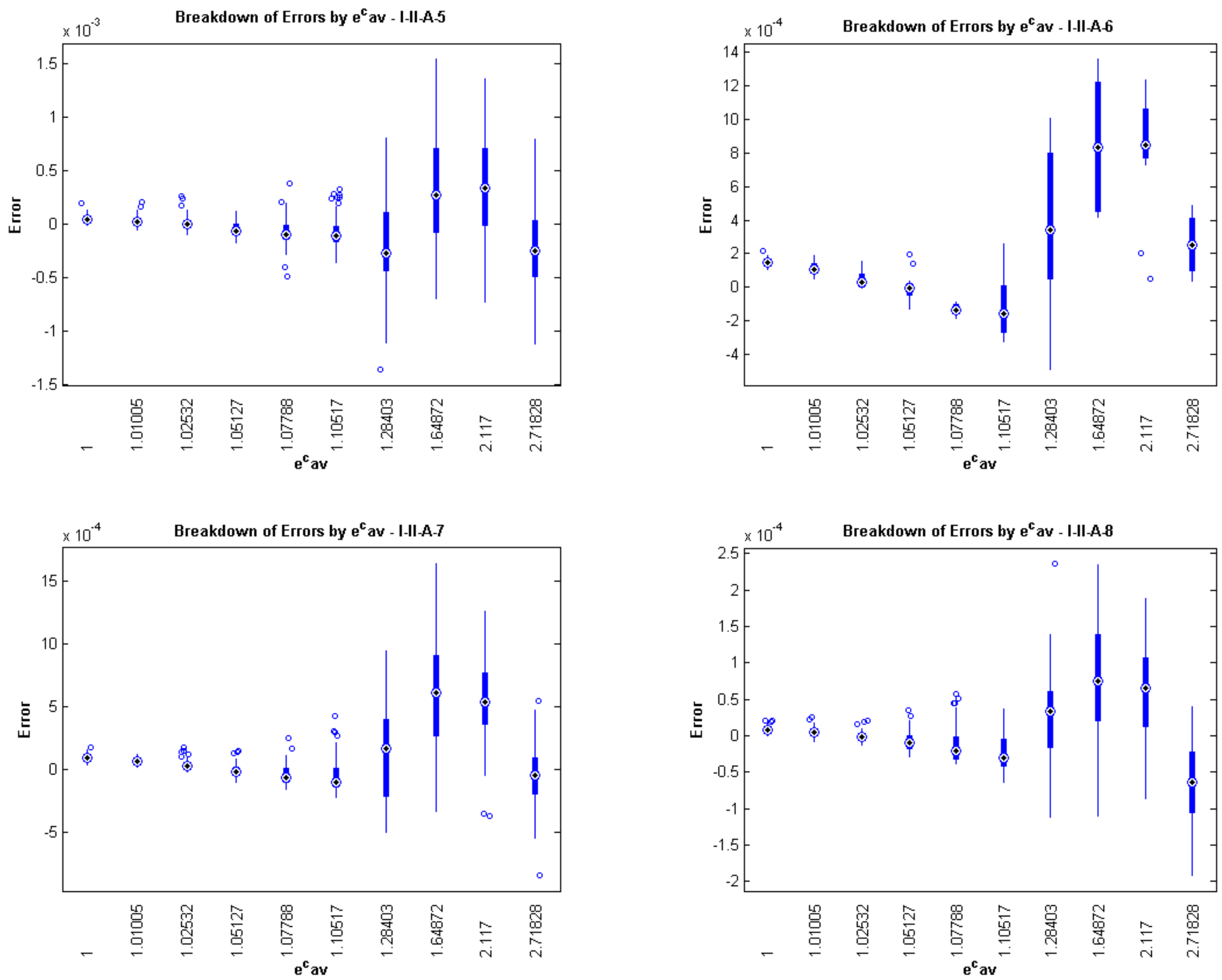


Figure J.9: Breakdown of Errors by $e^{c_{av}}$ using the Optimal Regression Model for all Data Sets (*cont.*)

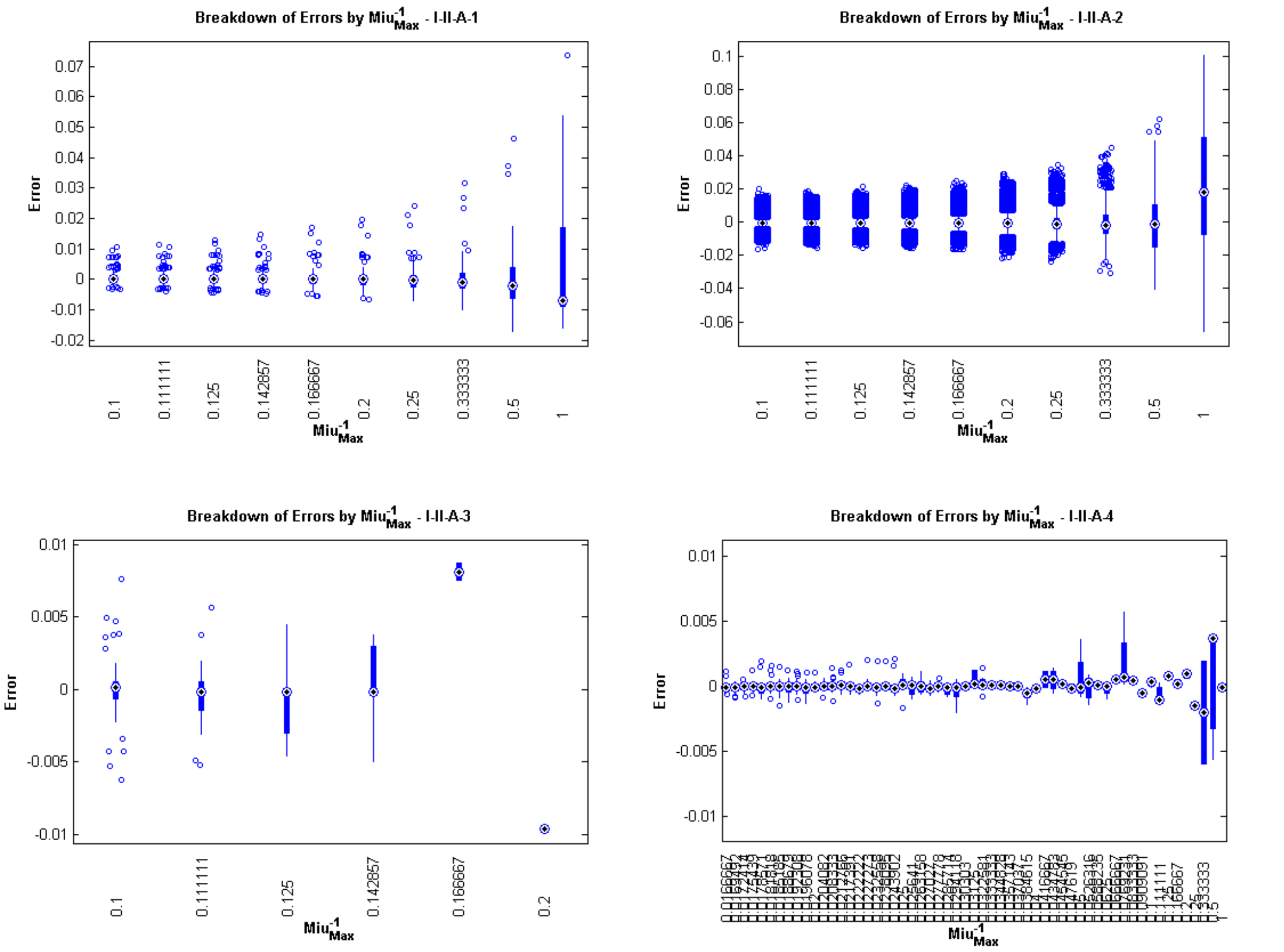


Figure J.10: Breakdown of Errors by $\mu_{\max}{}^{-1}$ using the Optimal Regression Model for all Data Sets

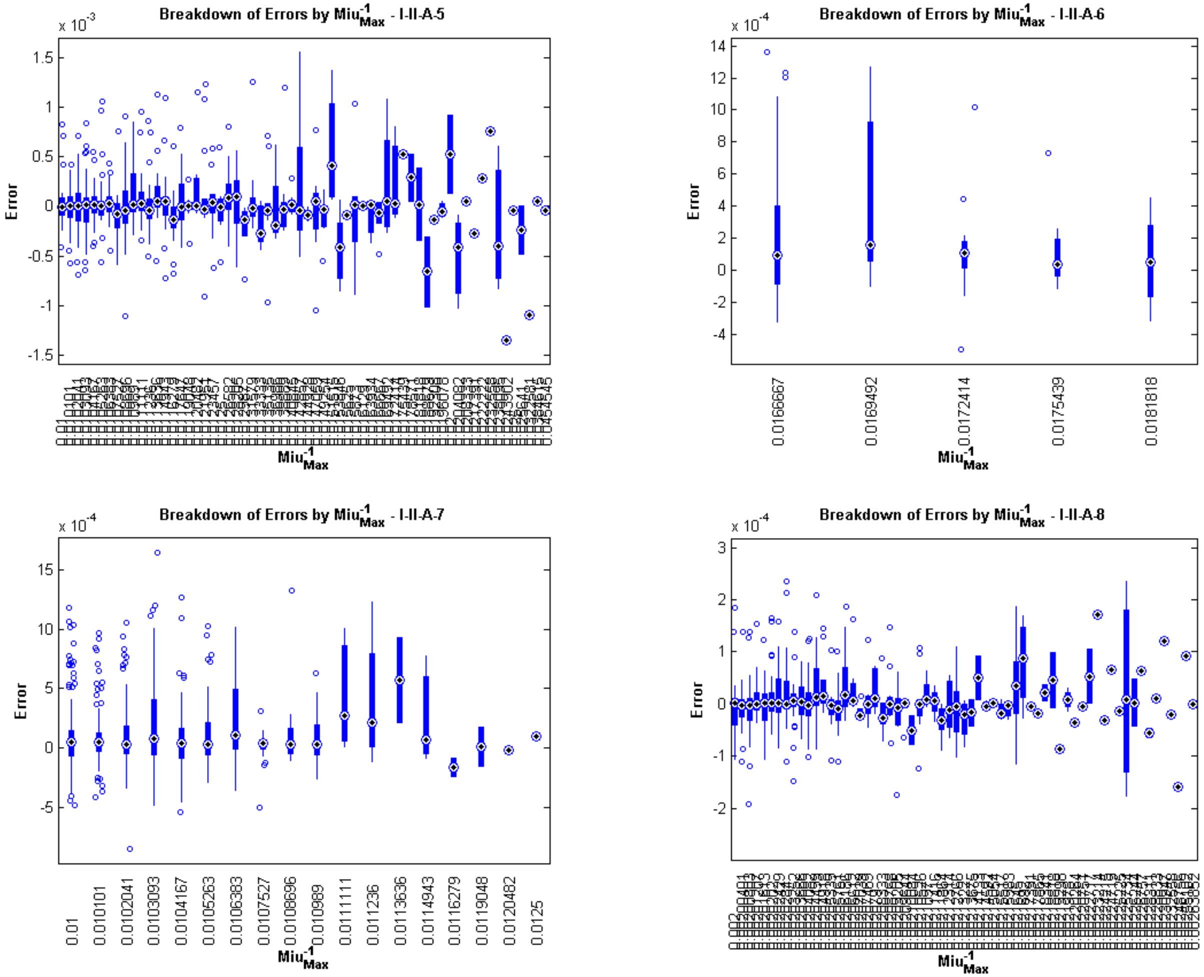


Figure J.10: Breakdown of Errors by μ_{max}^{-1} using the Optimal Regression Model for all Data Sets (*cont.*)

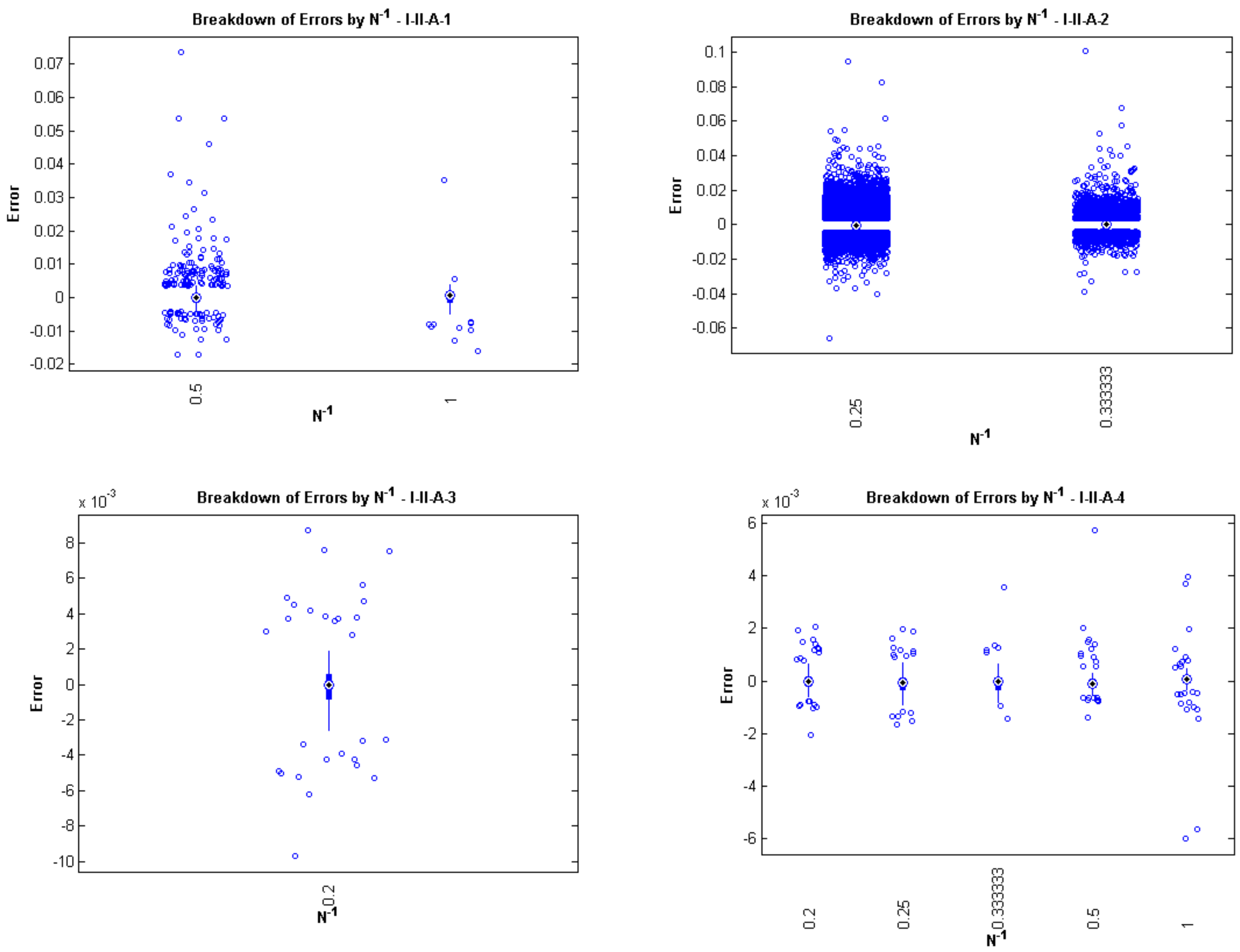


Figure J.11: Breakdown of Errors by N^{-1} using the Optimal Regression Model for all Data Sets

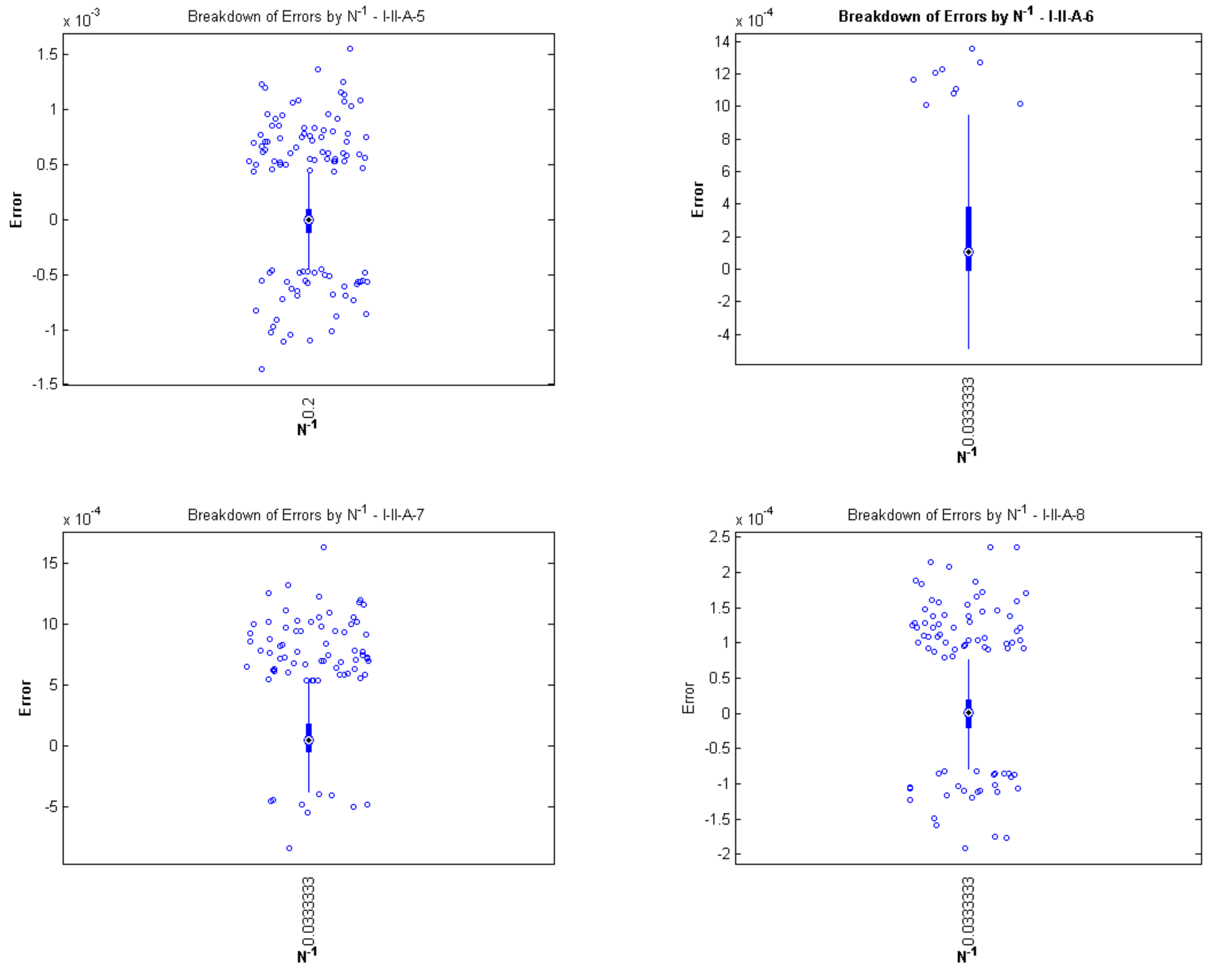


Figure J.11: Breakdown of Errors by N^{-1} using the Optimal Regression Model for all Data Sets (*cont.*)

APPENDIX Q – Autonomous Control Framework – Optimum Delivery Schedule for Decision Combination Scenario Number 1 of the Real-world Case Study

Refer to Section 7.3

Table Q.1: Optimum Delivery Schedule for Decision Combination Scenario Number 1 of the Real-world Case Study

Time	P1						P2							Total								
	Batch and Load (In)		Arrival on Site		Discharged		Batch and Load (In)		Arrival on Site		Discharged			Batch and Load (In)			Arrival on Site			Discharged		
	6m3	8m3	6m3	8m3	6m3	8m3	6m3	8m3	6m3	8m3	6m3	8m3		6m3	8m3	Total	6m3	8m3	Total	6m3	8m3	Total
07:00	0	0	0	0	0	0	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0
07:30	0	1	0	0	0	0	0	0	0	0	0	0		0	1	1	0	0	0	0	0	0
08:00	0	2	0	1	0	0	0	0	0	0	0	0		0	2	2	0	1	1	0	0	0
08:30	0	2	0	1	0	0	0	0	0	0	0	0		0	2	2	0	1	1	0	0	0
09:00	0	3	0	1	0	1	0	0	0	0	0	0		0	3	3	0	1	1	0	1	1
09:30	0	4	0	2	0	1	0	0	0	0	0	0		0	4	4	0	2	2	0	1	1
10:00	0	4	0	3	0	2	0	0	0	0	0	0		0	4	4	0	3	3	0	2	2
10:30	0	5	0	3	0	2	0	0	0	0	0	0		0	5	5	0	3	3	0	2	2
11:00	0	5	0	4	0	3	0	0	0	0	0	0		0	5	5	0	4	4	0	3	3
11:30	0	6	0	5	0	3	0	0	0	0	0	0		0	6	6	0	5	5	0	3	3
12:00	0	7	0	5	0	4	0	0	0	0	0	0		0	7	7	0	5	5	0	4	4
12:30	0	7	0	6	0	5	0	0	0	0	0	0		0	7	7	0	6	6	0	5	5
13:00	0	8	0	7	0	5	0	0	0	0	0	0		0	8	8	0	7	7	0	5	5
13:30	0	8	0	7	0	5	0	0	0	0	0	0		0	8	8	0	7	7	0	5	5
14:00	0	9	0	8	0	7	0	0	0	0	0	0		0	9	9	0	8	8	0	7	7
14:30	0	10	0	9	0	7	0	0	0	0	0	0		0	10	10	0	9	9	0	7	7
15:00	0	10	0	9	0	8	0	0	0	0	0	0		0	10	10	0	9	9	0	8	8
15:30	0	11	0	10	0	9	0	0	0	0	0	0		0	11	11	0	10	10	0	9	9
16:00	0	12	0	10	0	9	0	0	0	0	0	0		0	12	12	0	10	10	0	9	9
16:30	0	12	0	12	0	10	0	0	0	0	0	0		0	12	12	0	12	12	0	10	10
17:00	0	13	0	12	0	10	0	0	0	0	0	0		0	13	13	0	12	12	0	10	10
Total	**0**	**13**	**0**	**12**	**0**	**10**	**0**	**0**	**0**	**0**	**0**	**0**	**Total**	**0**	**13**	**13**	**0**	**12**	**12**	**0**	**10**	**10**

APPENDIX M – Autonomous Control Framework – Chosen Best Operational Settings for the Autonomous Control and Optimisation Methods with the 16 Decision Combination Scenarios of the Real-world Case Study

Refer to Section *7.3*

Table M.1: Chosen Best Operational Settings for the Developed Formula-based Autonomous Control Method with the 16 Decision Combination Scenarios of the Real-world Case Study

Experiment Number	No of Deliveries from the Same Concrete Plant	Concrete Plant	Load Size (m3)	Multiple Concrete Plant Usage	Arrival Rate (1/min)		Mean Site Delay (min)
					Mean	Standard Deviation	
E1	1	2	8	S	49	0	0
E2	1	2	8	S	49	0	0
E3	1	1	8	S	35	0	0
E4	1	2	8	S	49	0	0
E5	1	2	8	S	49	0	0
E6	1	1	8	S	35	0	0
E7	1	2	8	S	49	0	0
E8	1	2	8	S	49	0	0
E9	1	1	8	S	35	0	0
E10	1	1	8	S	35	0	0
E11	1	2	8	S	49	0	0
E12	1	1	8	S	35	0	0
E13	1	1	8	S	35	0	0
E14	1	2	8	S	49	0	0
E15	1	1	8	S	35	0	0
E16	1	1	8	S	35	0	0

Decision Variable

Table M.2: Chosen Best Operational Settings for the Developed Hybrid Autonomous Control Method with the 16 Decision Combination Scenarios of the Real-world Case Study

Experiment Number	No of Deliveries from the Same Concrete Plant	Concrete Plant	Load Size (m3)	Multiple Concrete Plant Usage	Arrival Rate (1/min)		Mean Site Delay (min)
					Mean	**Standard Deviation**	
E1	1	2	8	S	49	0	0
E2	1	2	8	S	49	0	0
E3	1	1	8	M	42	0	0
E4	2	2	8	S	25	0	0
E5	1	2	8	S	49	0	0
E6	1	1	8	M	42	0	0
E7	2	2	8	S	25	0	0
E8	1	2	8	S	49	0	0
E9	2	1	8	M	49	0	0
E10	1	1	8	M	42	0	0
E11	2	2	8	S	25	0	0
E12	2	1	8	M	49	0	0
E13	1	1	8	M	42	0	0
E14	2	2	8	S	25	0	0
E15	2	1	8	M	49	0	0
E16	2	1	8	M	49	0	0

Decision Variable

Table M.3: Chosen Best Operational Settings for the Simulation-based Optimisation Method with the 16 Decision Combination Scenarios of the Real-world Case Study

Experiment Number	No of Deliveries from the Same Concrete Plant	Concrete Plant	Load Size (m3)	Multiple Concrete Plant Usage	Arrival Rate (1/min)		Mean Site Delay (min)
					Mean	Standard Deviation	
E1	1	2	8	S	53	1.8	0
E2	1	2	7	S	62	7.3	0
E3	1	1	8	S	48	3.1	0
E4	2	2	8	S	40	2.9	0
E5	1	2	8	S	56	0	9
E6	1	1	8	S	54	5.4	0
E7	2	2	7.5	S	40	4.8	0
E8	1	2	8	S	66	13.4	9
E9	1	1	8	S	63	6.3	0
E10	1	1	8	M	52	7.6	9
E11	1	2	8	S	50	1.9	2
E12	1	1	8	M	66	9.6	0
E13	1	2	7.5	S	54	4.7	2
E14	1	2	8	S	28	2.6	0
E15	1	1	8	S	64	0	10
E16	1	1	6	S	53	0.5	8

Decision Variable