College of Engineering, Design and Physical Sciences

PhD Systems Engineering

Title:

Modelling and Design of the Eco-System of Causality for Real-Time Systems

Author:

MORAD DANISHVAR

Supervised by:

Dr ALI MOUSAVI

Professor JOHN STONHAM

Date: March 2015

A thesis submitted in fulfilment of the requirement for the award of the degree of Doctor of Philosophy in Systems Engineering

ABSTRACT

The purpose of this research work is to propose an improved method for real-time sensitivity analysis (SA) applicable to large-scale complex systems. Borrowed from the EventTracker principle of the interrelation of causal events, it deploys the Rank Order Clustering (ROC) method to automatically group every relevant system input to parameters that represent the system state (i.e. output). The fundamental principle of event modelling is that the state of a given system is a function of every acquirable piece of knowledge or data (input) of events that occur within the system and its wider operational environment unless proven otherwise. It therefore strives to build the theoretical and practical foundation for the engineering of input data. The event modelling platform proposed attempts to filter unwanted data, and more importantly, include information that was thought to be irrelevant at the outset of the design process. The underpinning logic of the proposed Event Clustering technique (EventiC) is to build causal relationship between the events that trigger the inputs and outputs of the system. EventiC groups inputs with relevant corresponding outputs and measures the impact of each input variable on the output variables in short spans of time (relative real-time). It is believed that this grouping of relevant input-output event data by order of its importance in real-time is the key contribution to knowledge in this subject area. Our motivation is that components of current complex and organised systems are capable of generating and sharing information within their network of interrelated devices and systems. In addition to being an intelligent recorder of events, EventiC could also be a platform for preliminary data and knowledge construction. This improvement in the quality, and at times the quantity of input data, may lead to improved higher level mathematical formalism. It is hoped that better models will translate into superior controls and decision making. It is therefore believed that the projected outcome of this research work can be used to predict, stabilize (control), and optimize (operational research) the work of complex systems in the shortest possible time.

For proof of concept, EventiC was designed using the MATLAB package and implemented using real-time data from the monitoring and control system of a typical cement manufacturing plant. The purpose for this deployment was to test and validate the concept, and to demonstrate whether the clusters of input data and their levels of importance against system performance indicators could be approved by industry experts. EventiC was used as an input variable selection tool for improving the existing fuzzy controller of the plant. Finally, EventiC was compared with its predecessor EventTracker using the same case study. The results revealed improvements in both computational efficiency and the quality of input variable selection.

ACKNOWLEDGEMENTS

I would like to acknowledge the contributions of the following individuals and groups for their help and support throughout the development of this thesis:

- My supervisors Dr Alireza Mousavi and Professor John Stoneham who have guided me with enthusiasm, support and patience through every aspect of this thesis, from conception to completion.
- Members of the Systems Engineering Research Group (SERG) at the College of Engineering, Design and Physical Sciences, Brunel University who provided an inspiring atmosphere of passion and joy in the workplace.
- Pedro Sousa and all the staff at Acontrol Automation Company, Portugal, who provided my experiment data and supported me during my research.
- Dr Rui Araújo from the Institute of Systems and Robotics, University of Coimbra, who invited me to work alongside his research team.
- The administrative staff at the College of Engineering, Design and Physical Sciences, Brunel University, without whose aid and support this research could not have been completed.
- I would like to express my sincere thanks to all of my friends who have encouraged and supported me during the last three years.
- Finally, and in time-honoured tradition, my biggest debt is acknowledged last. I would like to offer my deepest thanks to my family who have supported me in every conceivable way.

DEDICATION

I dedicate my Ph.D. thesis to my late father. When he was alive, he encouraged me to forge ahead until the very end. Following his death, his memory has pushed me to be stronger. May his soul rest in peace.

TABLE OF CONTENTS

ABSTRACT....ii
ACKNOWLEDGEMENTS.... iii
DEDICATION....iv
LIST OF FIGURES ix
LIST OF TABLES....xi
LIST OF ABBREVIATIONS.... xiii
1. Introduction....1
1.1. Identification of a gap 1
1.2. Research Hypothesis and Key Research Questions....4
1.2.1. Research Clarification....5
1.2.2. The Descriptive study I 5
1.2.3. Prescriptive study....6
1.2.4. Descriptive study II....6
1.3. The aim and objective of the thesis....7
1.4. Design a research methodology and thesis framework....10
1.4.1. The literature review theme....10
1.4.2. The methodology theme....12
1.4.3. The validation and conclusion theme....12
2. Systems Theory: Its Origins and Foundations....13
2.1. Origins of systems theory 13
2.1.1. Cybernetics....14
2.1.2. The fuzzy concept of information....15
2.2. Definition of system....15
2.2.1. Reduction to components....17
2.2.2. The system approach....18
2.2.3. Reduction to dynamics....18
2.2.4. System and environment....18
2.3. From systems theory to complex systems 19
2.3.1. Complex versus complicated systems....21
2.3.2....22
2.3.3. Simple versus comprehensive model....22
2.3.4. Modularity in technology....23
2.4. Basic system definitions 25
2.4.1. System model....25
2.4.2. Static and dynamic systems 25
2.4.3. Time-varying and time-invariant systems....26
2.4.4. Linear and nonlinear systems....26

2.4.5. Continuous -state and discrete-state systems 26
2.4.6. Deterministic and stochastic systems 26
2.4.7. Discrete-time and continuous-time systems............................26
2.4.8. The concept of events..27
2.4.9. Response times...28
2.4.10. Real-time systems 28

2.5. The concept of system control 28
2.6. The concept of feedback 29

2.6.1. Open-Loop and Closed-Loop systems 30

2.7. Summary of systems theory...30

3. A Review of the Input Variable Selection (IVS) 32

3.1. IVS and feature selection 33

3.1.1. Regression 34

3.2. Sensitivity analysis (SA)...35

3.2.1. The analytical relationship between input and output data 36
3.2.2. The statistical distribution of input variables 36
3.2.3. The computational overhead 37
3.2.4. Analytical methods...38
3.2.5. Sampling based methods 40
3.2.6. Heuristic based methods..47

3.3. Conclusion 48

4. Event Clustering Data Grouping Technique (EventiC) 49

4.1. The basic concept of Event Clustering 50
4.2. Data clustering methods for big data 52

4.2.1. Clustering algorithm categories 53
4.2.2. Criterion to benchmark clustering methods 54

4.3. Rank Order Clustering (ROC) technique..56

4.3.1. The ROC algorithm..57

4.4. EventiC method's basic parameters..59

4.4.1. Discrete Event Systems...60
4.4.2. Trigger Data and Event Data..60
4.4.3. Trigger Threshold (TT) 60
4.4.4. Event Threshold (ET)..60
4.4.5. Actual Value of the Data (AD) 60

4.5. Event clustering method's algorithm 60

4.5.1. The assumption of the proposed method................................61
4.5.2. Assumption 1: Delays 61
4.5.3. Assumption 2: Thresholds..61

4.5.4. Assumption 3: Homogeneity of the data series....................... 61
4.5.5. Trigger-event detection 62
4.5.6. Sample scan size.. 63
4.5.7. An example to understand the implementation of the ROC algorithm 63
4.5.8. Average sensitivity analysis (SA) weight 65
4.5.9. Cut-off threshold 66
4.6. Event clustering sensitivity analysis design process............................ 67
4.6.1. The computational structure of EventiC sensitivity method ... 68
4.7. Conclusion on Event clustering sensitivity analysis 69
5. An Industrial Case Study 70
5.1. The cement production selection as a case study for EventiC application and other alternatives 70
5.2. The cement production industry 72
5.2.1. Cement kiln process 73
5.2.2. Kiln system.. 74
5.2.3. The kiln's key performance factors and indicators.................. 76
5.3. A case study for EventiC application... 81
5.3.1. 5.2.1. Data collection and pre-processing 82
5.3.2. Sampling frequency.. 82
5.3.3. Implementation of EventiC on kiln operation......................... 83
5.4. EventiC and real-time plant control 87
5.4.1. Cement quality as a key performance indicator 92
5.4.2. Environmental impact as a key performance indicator 93
5.5. The detection of unknown factors affecting the behaviour of a system 93
5.6. Visualisation of input/output correlation with EventiC 95
5.7. EventiC application implementation.. 96
5.7.1. MATLBAB 96
5.7.2. DataBridge industrial environment 96
5.7.3. EventiC algorithm implementation 97
5.8. A Summary to the EventiC applications case study 98
6. An Application of EventiC in Input Variable Selection: A case study for Fuzzy Controller 100
6.1. Fuzzy systems 101
6.1.1. Fuzzy system concept.. 102
6.2. Fuzzy system controller 107
6.2.1. Types of Fuzzy Controller... 108

6.2.2. Design of membership function 108
6.2.3. Fuzzy rule reduction 111
6.3. Fuzzy controller in the cement industry 111
6.3.1. Cement production process 112
6.3.2. Select effective input and output variables in the cement quality control 114
6.4. EventiC integration to cement fuzzy controllers 117
6.4.1. Input variables delay 117
6.4.2. EventiC application 118
6.4.3. EventiC/fuzzy controller structure 119
6.4.4. Controller scheme design 122
6.4.4.1. Defining the controller's layers 122
6.4.4.3. Defining the membership function 122
6.5. Summary and conclusion 126
7. Assessment of the Efficiency and Validity of the EventiC 128
7.1. EventTracker sensitivity analysis technique 128
7.1. The EventTracker algorithm 130
7.2. A case study for comparing EventTracker and EventiC techniques .. 130
7.3. Conclusion 133
8. Conclusion, contribution, and future work 134
8.1. Contribution to knowledge 137
8.2. Future work 139
8.2.1. Research which bounds to EventiC technology: 139
8.2.2. Impact of the research 139
8.3. Overall benefits of the research outcome 140
REFERENCES 141
Appendices 158
Appendix A: Schematic of our industrial partner's SCADA output including the list of sensor names 158
Appendix B: The output of EventiC for 196 inputs over production rate (no. 137). In this example, the rate of data acquisition was 1 sample at every minute for a complete month. 170
.... 171
Appendix C: SCADA output for the kiln's 196 sensors in the first 30 samples (first 30 minutes) 172

LIST OF FIGURES

Figure 1.1. Current design project, problem solving, performance measurement and optimization process 4
Figure 1.2. Design project, problem solving, performance measurement and optimization processes with EventiC 9
Figure 1.3. Schematic flow of the thesis 11
Figure 2.1. Basic system concept 16
Figure 2.2. Simple modelling process 25
Figure 2.3. Continuous-time and discrete-time sample paths 27
Figure 2.4. Comparison of sample paths for Continuous-Variable Dynamic Systems (CVDS) and Discrete Event Systems (DES) 28
Figure 2.5. State space modelling with control input 29
Figure 2.6. Open-loop and closed-loop systems 30
Figure 3.1. Different SA methodologies 38
Figure 4.1. A system and controller in closed-loop 51
Figure 4.2. An overview of clustering taxonomy 53
Figure 4.3. Machine-part incident matrix 57
Figure 4.4. Incident matrix row ranking 58
Figure 4.5. Incident matrix column raking 59
Figure 4.6. Block diagonal structure matrix 59
Figure 4.7. Trigger-event Detection functionality on each time scale 62
Figure 4.8. Trigger event detection functionality over whole sampling time 63
Figure 4.9. Actual value of an example Input/output at t1 and t2 64
Figure 4.10. Build TDs and EDs from actual value in table 4.2 64
Figure 4.11. TDs and EDs incident matrix 65
Figure 4.12. Final diagonal ROC matrix 65
Figure 4.13. Averaged SA weight of an example TDs /EDs 66
Figure 4.14. Average SA weight of the example with CT=60% 66
Figure 4.15. Overall structure of sensitivity analysis method 68
Figure 4.16. Overall structure of EventiC method 69
Figure 5.1. Cement production overview 73
Figure 5.2. The kiln system diagram 74
Figure 5.3. KPIs of the customization factor in a manufacturing system 77

Figure 5.4. Calculation of energy consumption..79
Figure 5.5. The kiln's production rate SA with respect to 196 inputs over 1 month sampling snapshots with 90% cut-off threshold..84
Figure 5.6. Percentage of filtered TDs per CT and the ratio of false negatives over the production rate. 84
Figure 5.7. The kiln's CO_2 emission SA with respect to 196 inputs over 1 month sampling snapshots. 86
Figure 5.8. The kiln's temperature SA with respect to 196 inputs over 1 month sampling snapshots. 87
Figure 5.9. System state in time...88
Figure 5.10. Actual value of production rate and its maximum points over 1 month sampling snapshots. 89
Figure 5.11. Input/output relationship in current systems 94
Figure 5.12. Input / Output interrelationship via EventiC 94
Figure 5.13. Data decomposition and composition in EventiC 95
Figure 5.14. The output of EventiC with respect to 196 inputs over 43,000 sampling snapshots with 1% (0.01) sensors threshold..98
Figure 6.1. Range of logical values in Boolean (left) and fuzzy logic (right) 103
Figure 6.2. General scheme of a Fuzzy Controller (FC)...103
Figure 6.3. Examples of membership functions: a) Trapezoidal b) Triangular c) Gaussian .104
Figure 6.4. Gaussian Membership function..105
Figure 6.5. A rotary kiln plant 113
Figure 6.6. Cement quality key performance indicators and relevant inputs 115
Figure 6.7. EventiC/fuzzy controller integrated structure 119
Figure 6.8. Segregated EventiC/fuzzy controllers integrated structure 121
Figure 6.9. The membership functions for error of O2variables 123
Figure 6.10.The membership functions for error of CO variables 123
Figure 6.11.The membership functions for error of KA variables 123
Figure 6.12. The membership functions for error of BT controlled variable 124
Figure 6.13. The membership functions for error of APcontrol variable 125
Figure 6.14. The EventiC/fuzzy controller structure 127
Figure 7.1. General view of the EventTracker method for sensitivity analysis.....................129
Figure 7.2. EventiC and EventTracker CT versus percent of filtered TDs............................132

LIST OF TABLES

Table 2.1. Reasons why systems might be considered to be complex 22
Table 4.1. Comparison between five clustering method with respect to 3Vs..........................55
Table 4.2. Exclusive-Nor Functionality..61
Table 4.3. EventiC clustering method with respect to 3Vs...68
Table 5.1. Type of cement quality VS kiln temperature (Yao*et al*., 2010) 81
Table 5.2. EventiC's design requirement and design process..82
Table 5.3. Averaged SA weight of a selected number of the kiln's input data over the kiln's production rate 85
Table 5.4. Averaged SA weight of some selective kiln's input data over CO_2emission.........86
Table 5.5. Averaged SA weight of some selective kiln input data over kiln temperature.......87
Table 5.6. Five alternative clusters of input variables 90
Table 5.7. Five alternative clusters of input variables in minimum kiln's energy consumption 91
Table 5.8. Five alternative clusters of input variables in minimum kiln's CO_2 emission 91
Table 5.9. Five alternative clusters of input variables for medium quality of cement 92
Table 6.1. Limitations of conventional controllers versus the benefits of fuzzy controllers.102
Table 6.2. Formula used in t-norm (intersection) 106
Table 6.3. Formula used in implication operator..106
Table 6.4. Aggregation formula..107
Table 6.5. Kiln inputs from cement quality perspective...116
Table 6.6. Output variables of the kiln 117
Table 6.7. Input variables delay estimation 117
Table 6.8. Categorization of output and input variables...118
Table 6.9. Categorization of the kiln's input and output variables with CT=90% 118
Table 6.10. Kiln's inputs relationships with O_2 and CO...119
Table 6.11. Kiln's inputs relationships with KA 120
Table 6.12. Kiln's inputs relationships with BT and PT...120
Table 6.13. Authorized ranges for O_2, CO and KA variables...121
Table 6.14. Selected ranges for BT and PT variables...121
Table 6.15. BT and PT mean and variance after 1 month sampling.....................................124
Table 6.16. Kiln control variables mean and variance after 1 month sampling 124

Table 7.1. Comparison of EventiC and Event Tracker's SA weight of a selected number of a kiln's input data over kiln production rate 131
Table 7.2. Comparison table between EventiC and EventTracker techniques 132

LIST OF ABBREVIATIONS

Abbreviation	Stands for
AD	Actual Value of the Data
ANN	Artificial Neural Network
ANOVA	Analysis Of Variance
CF	Cell Formation
CM	Cellular Manufacturing
COG	Center Of Gravity
CPS	Cyber Physical System
CS	Customer Satisfaction
CT	Cut-Off
CVDS	Continuous-Variable Dynamic Systems
DAQ	Data Acquisition
DES	Discrete Event System
ED	Event Data
EFC	EventiC-Fuzzy Controller
ET	Event Threshold
EDIM	Event-Driven Incidence Matrix
EF	Efficiency
FAST	Feature Selection
FS	Fourier Amplitude Sensitivity Test
FLE	Fuzzy Logic Engine
FLC	Fuzzy Logic Controller
FLS	Fuzzy Logic System
FORM	First Order Reliability Methods
FAST	Fourier Amplitude Sensitivity Test
GA	Genetic Algorithm
GT	Group Technology
GHG	Green House Gasses
HIL	Hardware In Loop
INCOSE	International Council on System Engineering
I/O	Input/output

IoT	Internet of Things
IVS	Input Variable Selection
IM	Inventory Management
KPF	Key Performance Factors
KPI	Key Performance Indicators
LHS	Latin Hypercube Sampling
MISO	Multi Input Single Output
MPIM	Machine-Part Incident Matrix
MMC	Modified Monte Carlo
OAT	One-At-a-Time
PCA	Principal Component Analysis
PLC	Programmable Logic Controller
PR	Productivity
ROC	Rank Order Clustering
RU	Resource Utilization
SA	Sensitivity Analysis
SORM	Second Order Reliability Methods
SI	Sensitivity Index
SCADA	Supervisory Control and Data Acquisition Systems
TD	Trigger Data
TT	Trigger Threshold

1. Introduction

The motivation for this research project is to take a short step towards a greater understanding of complex systems' behaviour. This step is facilitated by modern technological advancements and the ability to construct and develop complex systems which are not only able to generate large amounts of data, but also analyse the performance of such systems.

As technology improves further, Systems are becoming more complex as they are developed. Scientists can access more data and are looking for better ways to interpret and ultimately solve complex problems. However, with large data the interpretation becomes increasingly difficult. We borrow a quote from Lee Segall (Cited in Livingston &Antal, 2009): "*A man with one watch knows what time it is, a man with two watches is never quite sure".* It refers to the potential drawback of having too much contradictory information when making decisions.

A primary goal when observing complex systems is to describe the system as a function of the observed factors. Typically, a large number of factors are measured rather than a mere few. The main problem is no longer what to measure but rather how to interpret many measurements, knowing that some factors might not be contributing to system understanding. This approach can yield vast data sets, which are seldom readily interpretable. A model describes a transformation of the input parameters corresponding to observed factors to a response corresponding to system behaviour. Knowledge gained from variable selection is key stage of process optimisation.

1.1. Identification of a gap

"*In order to understand an organised whole we must know both the parts and the relations between them."*(Von Bertalanffy, 1972).

Man-made systems are designed and implemented such that they maintain their integrity when faced with volatile operational environments. The most challenging part of any engineering project is to ensure that the designed system functions correctly in meeting its specified functionalities for its given operating conditions. System complexity is increasing as systems are developed to meet ever demanding requirements or are required to possess the necessary intelligence and physical capability to maintain system integrity when faced with perturbations in the operating environment. Such perturbations arise from both known and unknown internal or external event sources.

Such increased complexity can be partly attributed to physical properties such as the composition of building materials, mechanical structure, electrical and electronic enablers, and electro-mechanical interfaces and partly to the knowledge constructs surrounding artefacts (data emission and acquisition), data interpretation (input/output), state analysis (behaviour), action-reactions (performance), and what-if scenarios (prediction and optimization). The combination of electro-mechanical properties enabled with embedded electronics links such devices with the larger internet of other components and systems and this eco-system of a multiple system interacting with another is fundamentally changing our understanding and approach to design system integrity and fidelity.

To provide an example, a cement manufacturing specification is engineered and verified with respect to a set of required user functionalities. These specifications are normally translated into a structural model (an assembly of electro mechanical components) that integrates with data and communication control architectures. The modus-operandi of an engineered artefact is ultimately predetermined by its designers. This act of predetermination imposes a static nature on the system design and its functional attestation. This static design methodology is logical, and in general, yields tangible results in a given time span (research, design and development lifecycle). The design blueprint defines the principle components, their relationships and the overall functionalities of the system or artefact. Such an approach allows the developer to hardwire the system to its specification. Control mechanisms may be embedded into the system to enable it to cope with potential or predicted variances that arise within the system itself or its operational environment. Such predictive monitoring and control are salient features of any modern complex system and enable it to be smarter, more autonomous, and better able to adapt to the environment in which it operates.

However, often, these embedded control subsystems are designed in such a way that even though they are part of a much larger and complex system, they do not necessarily share information with other sibling subsystems that comprise the whole, but instead act as isolated, closed units. This separation could very well be intentional to ensure system integrity, or could also be unintentional, resulting from a lack of forethought at the design phase. Regardless of intent, the lack of accurate knowledge regarding interrelated events and triggers may very well result in so-called malfunctions or non-responsiveness from the system. These occurrences are well-known and reported as undocumented malfunctioning resulting from unknown circumstances. For example, the cement manufacturing production industry can be designed within a series of well-defined operational conditions. Under certain circumstances, due to environmental or structural perturbations, a series of unspecified events occur; events that were unforeseen in the original design specification or by the operator, and result in the production refusing to operate as expected. This chain of events may very well arise from a combination of factors that includes production rate, health and safety issues or raw material quality - a series of factors that result in unresponsiveness and system malfunction. Such events are normally unanticipated by the operator and unforeseen by the designer.

The sole purpose of any in-depth analysis of the combined event chain, its precedence, interrelationships, and ultimately influence on the expected system behaviour is to gain an understanding of why such a chain of internal and/or external events (random or predicted) has resulted in unexpected system behaviour. Depending on the complexity of the system this analysis is laborious to reconstruct and requires the assembly of information from multiple sources that includes communication systems and data-mining exercises, all performed in conjunction with consultation and coordination with a number of specialist individuals/teams/institutions. This exercise can be time consuming and very expensive to undertake and often the results are inconclusive due to a lack of meaningful information or expert disagreement. The more inaccurate the data and knowledge available to the reconstruction, then the more subjective is any expert agreement or disagreement. The given analysis will result in actions such as design updates, new operational guidelines, adjustment to training procedures, and new safety and security protocols.

The conventional modelling and design of physical systems normally relies on a set of known linear/non-linear differential equations or analytical models to describe the physical behaviour of the systems that they are meant for. The knowledge about the input variables (i.e. excitation parameters) and their impact on performance (i.e. output) requires high levels of background knowledge, know-how and effort by highly-qualified experts. The shortening and improving of the process of input data selection and analysis would therefore have major economic benefits for the stakeholders.

Observations in the manufacturing, automotive and aerospace industry reveals that the process of evaluating and validating the correctness and accuracy of the models in real time is extremely time consuming and expensive. For example, the Hardware in-the-Loop (HiL) models are normally based on limitations imposed by modelling experts on the numbers and extent of input variables (often restricted to 'known known'). Most practitioners rely on trial and error and/or costly destructive and non-destructive testing. Figure 1.1 can be viewed as an interpretation of the current modelling process.

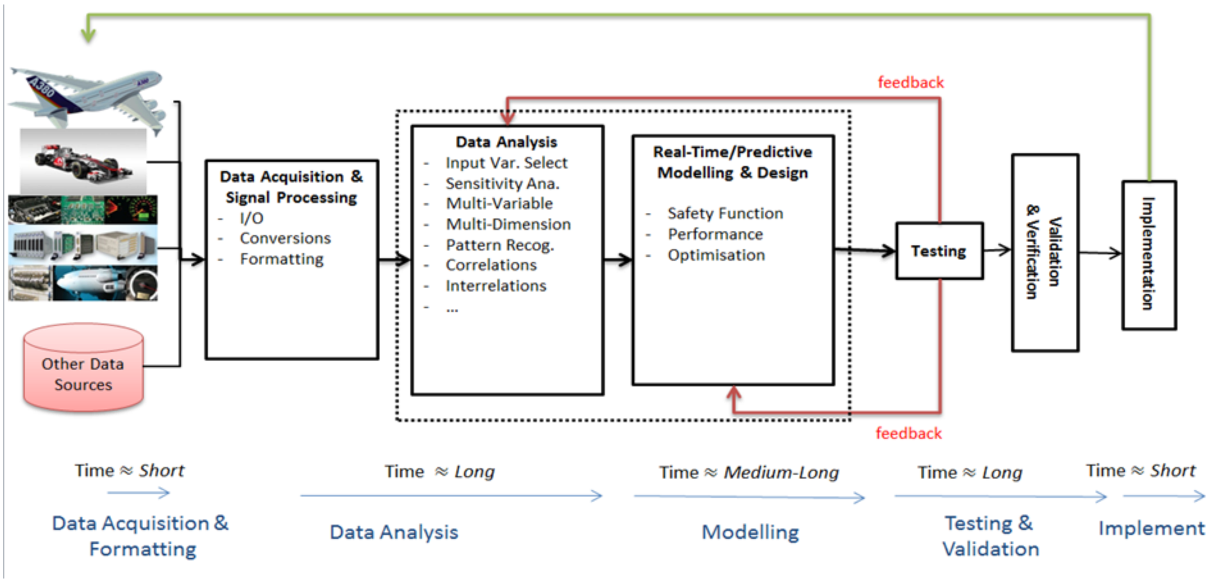


Figure 1.1. Current design project, problem solving, performance measurement and optimization process

1.2. Research Hypothesis and Key Research Questions

The Hypothesis of this research is that: *"all the available knowledge about the internal and external events surrounding a defined system has an effect on its state, unless proven otherwise."* Although this has been philosophically argued by Lorenzo (2000) discussed in the literature, but the key objective of this thesis is to provide an optimal practical and technical platform to prove the interrelationships between the measurable parameters representing systems state. A manufacturing process is then used as a case study to verify the hypothesis.

The motivation for the proposed research is that modern manufacturing systems are capable of sensing, communication, modelling, and providing corrective actions (actuation) to adjust to

changing operational and functional requirements. The flexibility to adapt can only be assured if data is succinctly interpreted and translated into corrective actions in a timely manner.
Within the context of the thesis, the key questions that this thesis intends to address are:

1. What would the implication of unbiased increase of input data and their potential relationships with one another and the system outputs be on our understanding of systems behaviour? Addressing the scalability problem.
2. How would the new knowledge about the new interrelationships between systems components allow us to better define performance (e.g. cost, quality, reliability, and fidelity)?
3. How would tracking and relating the events that represent the observable behaviour of the system lead to an increased insight to system functionalities and does it lead to more desirable improvements in the control of the system, and its optimal operation?

The four stages of design research methodology (DRM) proposed by Blessing and Chakrabarti (Blessing & Chakrabarti, 2009) is adopted to devise a framework for the planning and implementation of the research programme.

1.2.1. *Research Clarification*

The research clarification stage helps to clarify the current understanding and the overall research objectives, develop a research plan and provide a focus for the subsequent stages.
At the first instance the author posed the challenges and the fundamental research questions. In the following sections of this chapter the aim and objectives of the research are explained (section 1.3).

Chapters 2, 3, and parts of chapters 4 and 6 appraises the existing literature relevant to the subject area. The intention of the literature review was to identify the gaps in the existing technologies and methods of input variable selection, sensitivity analysis, complexity, systems modelling, as well as mathematical techniques for data management. The analysis of the state-of-the-art provides the focal point of the innovation required to bridge the gaps.

1.2.2. *The Descriptive study I*

Building from the literature analysis, an outline of the potential solution to the problem was emerges in chapter 4. In this chapter the author discusses the roadmap and the framework of the solution to be designed and applied. Based on the knowledge acquired about the definition of systems and in specific the definition of a manufacturing process system the components of the experimental design are established. Upon the establishment of the framework, a model of the real system is created in the laboratory environment. By establishing the emulator the key factors that influence the performance of the system can be tested and validated through observable experiments and feedback from experts in Cement Factory. The author acknowledges the support he received from AControl. The result of this stage would be a reference model (Calderon, 2010).

Through this stage the design architecture of the proposed solution emerges. The emerged Event-Clustering solution is the outcome of evolutionary improvements to existing data and knowledge engineering and systems modelling, with comparisons made with the existing solution in input variable selection, sensitivity analysis and big data clustering methods. The resultant model was then compared to the most relevant technique i.e. EventTracker.

1.2.3. *Prescriptive study*

Chapters 5 and 6 are dedicated to implementation of the preliminary design and conducting experiments in the laboratory and on the field to validate and verify the resultant Event-Clustering algorithm. Various tests were conducted to test the hypothesis and to establish whether the key research questions were answered. The tests included laboratory based simulation (discrete, deterministic and probabilistic) models, as well as filed data comparisons.

The models and the resulting Event Clustering (EventiC) method was implemented in a live operations of a cement factory. A novel fuzzy inference modelling for performance optimisation was introduced by utilising the results of Event Clustering.

1.2.4. *Descriptive study II*

In chapters 7 and 8 the author evaluates whether the key questions of the research were answered and whether the new solution provides a novel contribution to the existing body of knowledge. At this stage the efficacy, the usability and the applicability of the solution is evaluated.. The validation process examines whether the proposed EventiC has effectively answered the questions posed and has contributed to improving the quality of systems monitoring and control.

1.3. The aim and objective of the thesis

The aim of the this thesis is to propose a data and knowledge engineering platform to meet the challenges of the dynamic, autonomous, adaptive and self-organising embedded systems, and, seamless/secure interaction of the embedded system/cyber-physical systems with their environment. This research intends to interrelate the internal dynamics of the components embedded system or network of devices with that of other systems that create its operational environment and the first of its kind, the proposed technique will be able to evaluate the impact of each member of the cyber-physical system on the performance, stability and overall behaviour of the system in real-time.

There is a major difference between the proposed event modelling and the classical approach to data modelling and management. In the classical approach a state vector is expressed as a series of input data representing more complex information (e.g.$V_N = [x_1, \dots x_n]$).]). Any subsequence operation on the vector is based on the assumption that the vector is the true representation of the data series (a factual known). In contrast, the proposed event modelling (EventiC) technique does not make such an assumption about the input data. From its perspective there is no prior knowledge of any association between the total system input parameters (excitation parameters) and the output(s), thus it is 'unaware'. EventiC treats data as an 'unknown' mass that needs to be organised prior to any formal expression of the information.

Event modelling begins with event vector$E_S = [e_1, \dots e_s]$, where event E is expressed by all observable events that occurred at a given instant (S = sample space). By implementing the EventiC algorithm the problem statement becomes a more reliable expression of $E_N = [e_1, \dots e_n]$, where$e_1$toe_n, and $N \leq S$, input events are the true representation of E. By reinstating the actual values of the variables of the state vector, $V_n = [x_1, \dots x_n]$this could be considered as a more reliable state vector prepared for subsequent operations such as transfer functions, inferential models and other forms of data manipulation and knowledge representation.

The outcome of this research is to propose a platform that automates and integrates the process of data acquisition and analysis of raw data in near real-time would appear to be timely. As noted above, the acquisition process of large-scale data and organising it in the form of cross interrelationships and clusters of relevance, takes place at the lowest layer of interface between the physical system and the higher-level information framework. The proposed method could indeed be considered the linking point between the engineering of physical systems and higher-level data modelling constructs, thus allowing a true cyber interface of complex devices that can help each other to stabilise or function at optimal required outputs. This is achieved in two dimensions, system state change at defined intervals (vertical) and changes in time domain (horizontal), throughout the study span.

The first question posed in the research is answered by a combination of literature review (Chapters 2, 3 and partially chapters in 4), where the definitions of system and methods of knowledge engineering and management is presented. Furthermore, by designing the experiments through the systems simulation and observations on the live activities of the factory, the scalability of the approach was facilitated. Using live feed data from the factory and cross correlating the system inputs and outputs it was possible to assess the issues of significant increase in quantity and quality of data to monitor the behaviour of the system.

The second question posed in this research question is addressed in chapters 5 and 6, where the state of the system or the output parameters of the system is defined by the Key Performance Indicators specified in the manufacturing industry. These indicators are product quality, productivity, production efficiency, resource utilisation, and inventory. They are represented by well-established transfer functions defined in manufacturing systems literature. The raw data emitting from the sensors and actuators in the plant integrated by a SCADA system represent the input parameters of the system.

The designed experiment enabled the simulation of acquisition of real-time data from the plant and conducting sensitivity analysis against the event that takes place during the production process. The observations took over 30 days at a rate of 1 minute sampling rate.

The sensitivities of all KPIs were assessed against all 196 available input data series. Whilst originally this was not the case. Individual KPIs were only connected to a pre-specified set of input series. Thus enabling us to assess the efficacy of de-modularity of the system.

Finally to answer the third question of the thesis, a full case study was implemented with the involvement of the factory and production engineering staff. The new modulations and extracted relationships between systems input and output parameters lead to better understanding of the system behaviour and more over reducing the lead time to return the system to optimal performance. This was proven by connecting resultant EventiC to the factory's fuzzy controller, in which new membership functions showed to represent the dynamics of the system more effectively and accurately.

EventiC achieves this by simply (a) interpreting changes in the values of input-output (I/O) data as I/O events (b) detecting if the I/O events coincide, and (c) groups the I/O events, as related events. This process happens at a specified time interval called the scan rate. Scan rates can range from microseconds to seconds or minutes. Each scan registers a scenario of input-output, akin to recording a clip in a film. The weight of an input on an output is calculated using the basic logic of the number of coincidence in a time span. At the end, and for the purpose of modelling, we return to the actual values of the I/O data. Such an approach can be considered a novel one in the understanding of large-scale raw data.

With the help of a case study, we demonstrate the application of EventiC in assembling the necessary knowledge (data analysis) for the purpose of systems modelling and control optimization. The automation of the preliminary data analysis has significantly reduced time system modelling, design and validation. Figure 1.2 represents the new modelling platform. In this model EventiC sits between raw data acquisition platform and upper layer modelling and design platform as a middleware.

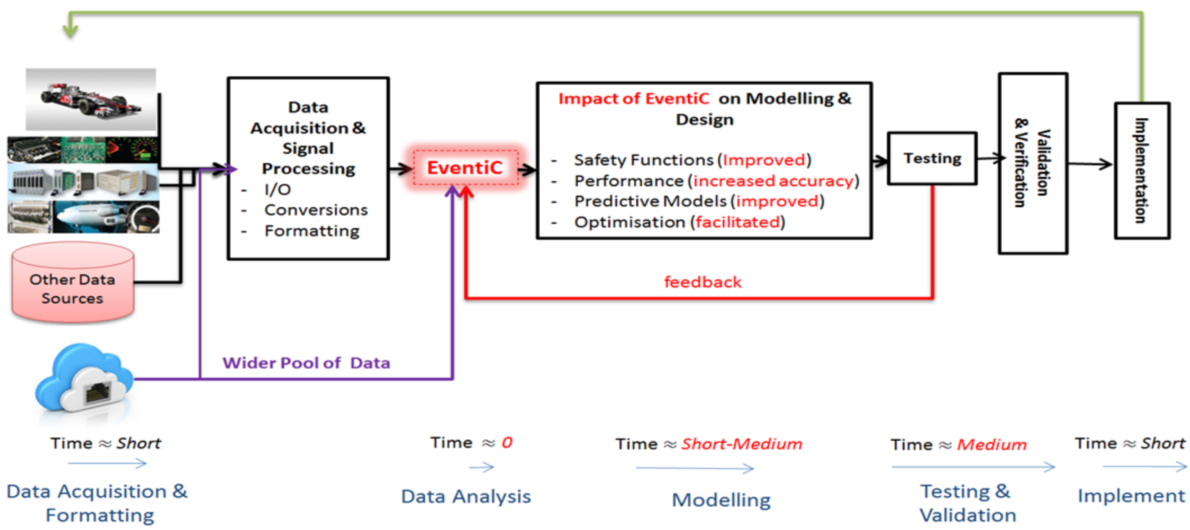


Figure 1.2. Design project, problem solving, performance measurement and optimization processes with EventiC

With expansion and fluctuation of modern complex systems, producing time-critical accurate knowledge about the state of the system remains and continues to remain a major challenge for researchers and practitioners. Thus building a powerful eco-system of causal events could be one of the most natural approaches to de-cluttering the complexity that arises from larger datasets. The challenge would therefore be to improve the current technology of de-clustering complexity in a specific time span of the system life cycle.

In order to address the challenge of conceptualising complexity this thesis suggests the concept of Event clustering as a platform for managing the interrelationships and internal and external dynamics of the components of a system. It intends to understand the causal relationships between the system and its operational environment as the system changes state and boundaries in different time spans.

1.4. Design a research methodology and thesis framework

The thesis and its constituent parts, as shown in figure 1.3, were prepared under three themes and eight chapters.

1.4.1. *The literature review theme*

This theme spans two chapters (chapters two and three) and two sections of chapters four and six in which gaps and omissions within the current theories, methodologies and technologies are discussed.

1.4.1.1. Chapter two

This chapter introduces and reviews the philosophical aspects and origins of systems theory, cybernetics, information theory and system boundaries from the 1950s onwards. Following this, some fundamental definitions in the system's field which assist in deep understanding the rest of thesis will be explained.

1.4.1.2. Chapter three

In this chapter the Input Variable Selection (IVS) as a methodology to help with the reduction of irrelevant inputs on a system's output is explained in detail. This is followed by a discussion on the shortcomings of existing sensitivity analysis (SA) methods in real-time systems.

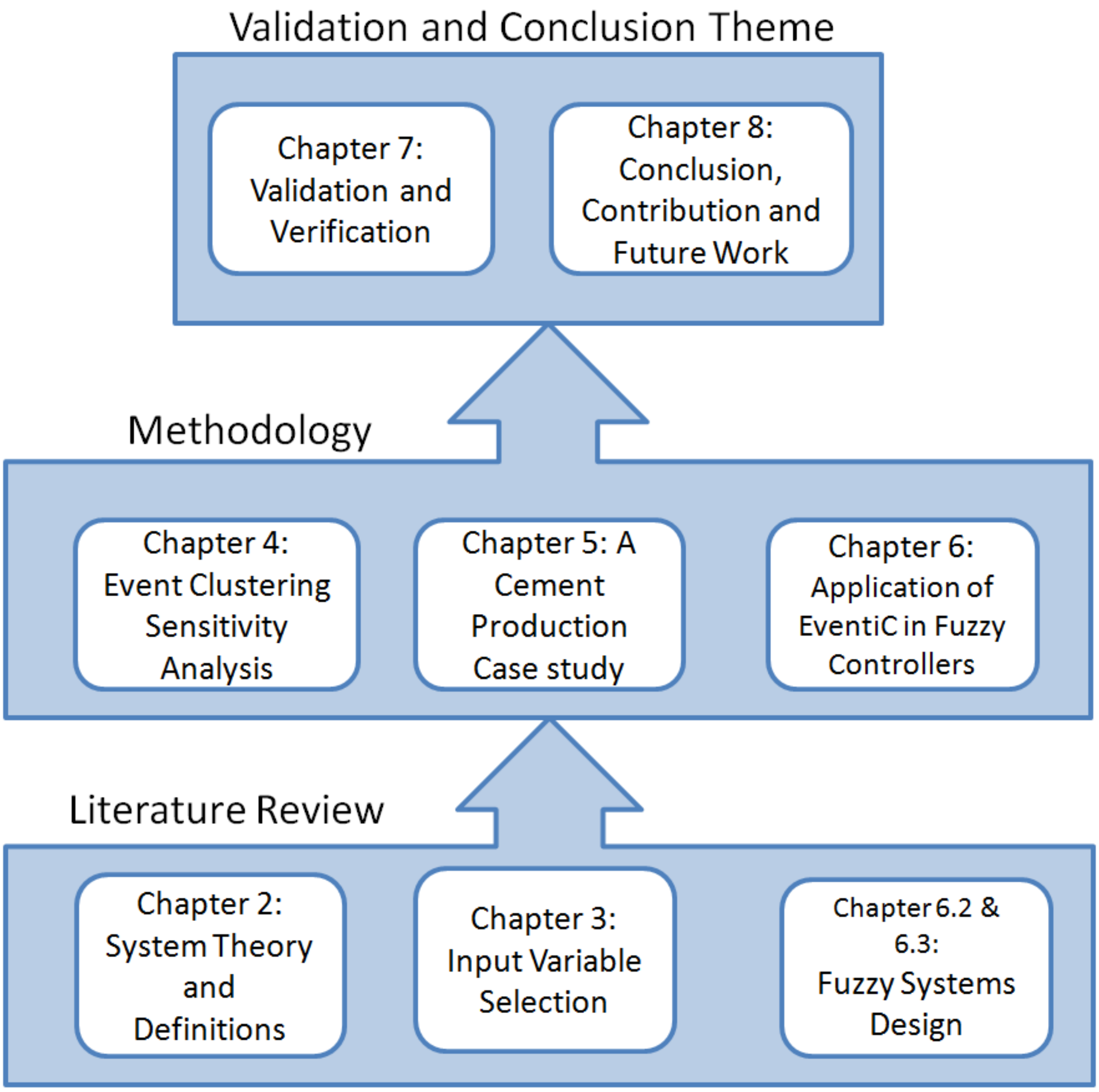


Figure 1.3. Schematic flow of the thesis

1.4.1.3. Chapter 4.2 and 6.2

In section 4.2 the existing clustering methods have been analysed and their strengths and weakness are reviewed. In section 6.2, literatures in Fuzzy controllers auto extraction rules has been explained.

These two chapters and two sections help to clarify the eco-system of complex systems definitions. They also identify the existing gap in SA methods and weakness of clustering and visualisation methods in dealing with big data in real-time.

1.4.2. *The methodology theme*

This theme comprises three chapters (chapters four, five and six) addressing the approach taken to provide solutions to the problems described in the previous chapters. The theme therefore covers the efficiency of the proposed sensitivity analysis technique, its application via an empirical study and the proposed method application in conjunction with fuzzy controllers.

1.4.2.1. Chapter four

This chapter introduces the new methodology of EventiC with reference to the manufacturing industry. The chapter details EventiC's improvements to the quality of real-time event recording by collecting, collating and classifying data gathered and generated within complex embedded devices/systems from internal and external events.

1.4.2.2. Chapter five

In this chapter the proposed research is implemented upon our industrial partner's raw data, obtained from the SCADA of their cement production plant. The EventiC application is applied to the SCADA outputs to control and optimize three main key performance factors in the cement industry: production rate; environmental impact; and product quality.

1.4.2.3. Chapter six

The proposed method application of EventiC in integration with fuzzy controllers is discussed in chapter six. The chapter reveals how the integrated EventiC/fuzzy controllers are able to automate fuzzy controllers in input variable selection and fuzzy rules/inference table design.

1.4.3. *The validation and conclusion theme*

This final theme introduces the validation and verification of the proposed method application in chapter seven with final conclusions offered in chapter eight.

1.4.3.1. Chapter seven

The penultimate chapter validates and verifies the proposed method by comparing the proposed sensitivity analysis methodology with its predecessor method, EventTracker. The efficiency of the proposed sensitivity analysis method and other differences between EventiC and EventTracker form an additional topic of discussion.

1.4.3.2. Chapter eight

The final chapter concludes the research work and discusses the potential for future work and development.

2. Systems Theory: Its Origins and Foundations

This chapter will review the origins of systems theory, cybernetics, and information theory from the 1950s onwards. Following this, the fundamental definitions used in systems theory will be defined, explained and explored. It is proposed that this account of systems theory will aid the author to further conceptualise what is meant by evolving system boundaries.

The objective of this chapter is to develop a deeper understanding of systems /complex systems theory and its inter-relationship with both sub-components and the wider environment. This understanding could help better define and possibly contribute to a new perspective about the definition of system boundary. It is believed that such a re-examination of current systems theory provides the potential for a critical exploration of a boundless, rather than bound, system framework regarding the complex relationships and interactions between events and perceptions, and the world that systems represent. This boundless, rather than bound, system framework enables the wider acquisition, exchange and process of data from any given physical or virtual system.

2.1. Origins of systems theory

Systems theory originated in the late 1920s in the work of the philosopher and biologist Ludwig Von Bertalanffy and was subsequently furthered by Ross Ashby in the 1960s (Ashby, 1964). In the early 1950s systems theory integrated with cybernetics, and in the late 1950s with the fuzzy set approach in engineering science. Von Bertalanffy considered to be the founder and principal author of general systems theory was both reacting against reductionism and also attempting to revive the unity of science. His proposal was for a unified principle of sciences with the organizational structure of the individual areas able to join together to form a whole.

Von Bertalanffy (1972) defined general systems theory as a universal science of wholeness. The meaning of the somewhat mystical expression, ‘The whole is more than the sum of its parts’ can be understood, in simple terms, to mean that the constitutive characteristics are not explainable only from the characteristics of the isolated parts. The characteristics of the complex, therefore, appear as new or emergent (Laszlo, 1972). Von Bertalanffy (1968) posited that a system is a complex of interacting elements, open to and interacting with their environments. In addition, such systems can acquire qualitatively new properties through emergence, thus they are in continual evolution. Systems are generally self-regulating (through feedback) and system thinking concerns itself with addressing both part-to-whole and whole-to-part thinking about making connections between various elements so that they fit together as a whole.

In 1948 cybernetics and information theory emerged from two publications, by Wiener and Shannon respectively *Cybernetics: or Control and Communication in the Animal and the Machine* (Wiener, 1948), and *A Mathematical Theory of Communication* (Shannon, 1948). Shannon's theory popularized later by Warren Weaver (Weaver, 1949) represents the beginning of the then so-called 'information theory'. Today, Shannon's name is associated almost exclusively with mathematical information and communication theory, with these terms often used and considered synonymously. Whilst Shannon focused on the measurement and coding of 'information' in general communication systems, the concept of 'information' led Wiener to the idea of feedback as a principal of control. Wiener developed a theory of feedback and governance operating across any simple or complex system. He understood feedback processes to be those of information manipulation and decision making (Seising, 2010).

In summary, this thesis suggests a need to return to the origins of cybernetics to re-define system cybernetics. A cybernetic system, in general terms, should be self-governing, intelligent and possess interrelationships and internal dynamics of the components within the overall eco-system of systems and their environment.

2.1.1. *Cybernetics*

Wiener used cybernetics as an umbrella term to define the study of control and communication in the animal and machine worlds. His research reached a climax in the interdisciplinary field of cybernetics with his examination 'time series' relating to a series of measurable events, recurring constantly or discretely. From this, Wiener developed his prediction theory, in which an operator called 'predictor' was applied in each case to the preceding element of the time series. In theory, this predictor corresponds to a mathematical calculation scheme; in practice, such as predicting the flight path of an enemy object it was realized by a technical apparatus (Wiener, 1949, cited in Seising, 2010, p. 8).

Cybernetics not only evolved as a separate field with specific theories and methodologies, but also created the foundation for spreading the concepts and ideas of the general principles of systems organization, control, and evolution to other research areas, such as decision making, machine learning, knowledge engineering, etc. (Arnold, Siemers & Adamson, 2014).

2.1.2. *The fuzzy concept of information*

Inspired by Wiener's Cybernetics, Shannon's and Weaver's *The Mathematical Theory of Communication*, Lotfi A. Zadeh (Zadeh, 1952) introduced a new approach to deal with the fundamental problem of systems communication. One of the problem that initiated Zadeh's thoughts about not precisely specified quantitative measures led to his famous theory of fuzzy set in 1965 (Zadeh, 1965).Due to his highly influential General System Theory, which was a revolution in its time, understanding regarding the behaviour of complex systems has evolved from a classic systems theory to modern theories, and substantial progress has been made in the development of highly complex systems in various subject areas. This theory will be introduced in detail and integrated with the proposed event modelling EventiC technique in chapter six.

2.2. Definition of system

The concept of 'system' serves to identify those manifestations of natural phenomena and processes that satisfy certain general conditions. In the broadest conception, 'system' can be understood as a complex of interacting components, together with the relationships among them, that permits or allows the identification of a boundary-maintaining entity or process.

Ackoff (2000) defined system identification as a set of two or more interrelated elements with the following properties:

- Each element has an effect on the functioning of the whole.
- Each element is affected by at least one other element in the system.
- All possible subgroups of elements also possess the first two properties.

INCOSE (2015) defines system:

- A set of interacting components-whether human-made, naturally-occurring, or a combination of both.

By replacing the concept of element with component, the definition of 'system' covers any kind of formal (e.g. mathematics), effective (e.g. imaginative), or existential (e.g. real-time) grouping. In each case, a 'whole' made up of interdependent components in interaction is identified as the system. In its most basic definition, a system is a group of interacting components that conserves some identifiable set of relations with the sum of components plus their relations (i.e. the system itself) conserving some identifiable set of relations to other entities (including other systems). Systems theory differentiates between system structure (the inner composition of a system) and behaviour (its outer presentation).

Figure 2.1 shows a black box. The external behaviour of a system is the relationship it imposes between its input and output time histories. The system's input-output behaviour includes the pairs of data records (input time segments paired with output time segments) gathered from a real system or model. The internal structure of a system includes its state and state transition mechanism plus the state to output mapping. Knowing the system structure allows us to analyse and/or simulate its behaviour (Zeigler, 1999).

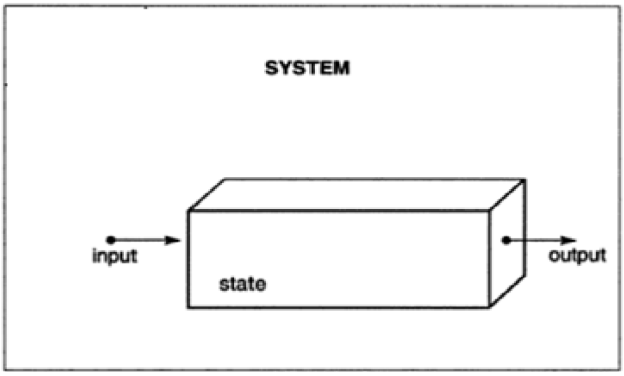


Figure 2.1. Basic system concept

System is one of those primitive concepts of which understanding might be left to intuition rather than an exact definition. Nonetheless, three representative definitions can be found in the literature on the subject (Cassandras &Lafortune, 2008):

- An aggregation or assemblage of things combined by either nature or man so as to form an integral or complex whole (Keating *et al.*, 2003).
- A regularly interacting or interdependent group of items forming a unified whole (Ramadge&Wonham, 1989).
- A combination of components that act together to perform a function not possible with any of the individual parts (Blanchard, 2004).

Boulding (1956) described General Systems Theory as a level of theoretical model building that lies between the generalized constructions of pure mathematics and the specific theories of specialized disciplines. Whilst mathematics attempts to organize highly generalized relationships into a coherent system, it is a system however, which does not have any necessary connections with the real world around us.

There are two prominent points to draw from these definitions. Firstly, a system is associated with a function which it is presumably intended to perform. It is worth indicating at this point that a system should not always be associated with physical objects and natural laws. For example, systems theory could be utilised as a framework for describing economic mechanisms, population dynamics or modelling human behaviour. Secondly, a system consists of interacting components.

In the next section, an important structural concept named 'reduction to components' will be explained. Questions such as 'How can a system be broken down into component systems?' and following on from this, the concept of composition, i.e., 'How can a system's components be coupled together to form a larger system?' could help to meet the research challenge of providing a dynamic and self-organising embedded system platform which has a secure interaction with its environment.

2.2.1. *Reduction to components*

The principal of heuristic innovation within the system approach is what may be called 'reduction to components' as practiced in the methodologies of classical science. Phenomena in the observed world are generally too complex to be understood by modelling all of their parts and interactions. Some form of simplification is necessary. Traditionally, a scientist simplifies complexity by dividing items into individual components and views such items in isolation from the complex set of relations that connect them with their environment, and ultimately, the rest of the world (Hooker, 1981).

The term 'reduction to components' has led to the accumulation of substantial information about specific entities and the interactions amongst them. 'Reduction to components' helps system analysts to learn how one component reacts to particular system stimulants/inputs and how each reacts to it part. However, this type of knowledge proves deficient in one important

respect for it does not disclose how complex things behave when exposed to a complex set of influences. In the real world almost every system contains a large number of components and is exposed to a large number of complex external events (Laszlo &Krippner, 1998).
In consequence, another heuristic become necessary, capable to simplifying complex system.

2.2.2. *The system approach*

Prior to the mid-twentieth century, definitions of 'system' were based upon dividing a 'whole' and all that it contained into smaller and distinct parts, predominantly for analysis and study in isolation (Cassandras & Lafortune, 2008). This approach to empirical phenomena is based on the belief that it is better to have specific detailed knowledge of smaller and more well-defined items than general knowledge of bigger and less well-defined ones. Therefore, attention is given to a system's parts regardless of their position, instead of a focus upon their interacting and integrated ensemble (Distefano et al., 2012).

In contrast to this, the 'system approach' considers the world in terms of an integrated system. It focuses on the 'whole system', instead of its components, as well as on the complex interrelationships amongst its constituent parts.

2.2.3. *Reduction to dynamics*

With regards to those systems that are too complex to be understood by modelling all of their parts and interactions, some form of simplification is necessary. The 'reduction to dynamics' method concentrates on the dynamics that define the characteristic properties, functions and relationships that are internal or external to the system. This method could be seen to cover the deficiencies of the 'reduction to components' approach.

In this method in contrast to 'reduction to components' which deconstructs a system to its smaller components and then analyses each component separately, investigation starts with identifying a system's stimulants /inputs and behaviours/outputs. The next step is to examine the impact of such stimulants on a system's behaviour. The key emphasis in this method is its focus on stimulant/behaviour (input/output) relationships and bonds, in addition to the examination of elements and components (Laszlo &Krippner, 1998).

2.2.4. *System and environment*

An important dynamic dependent feature relates to the effects of the external environment upon the system. In systems theory, the term 'environment' refers to those sets of objects whose change in attributes affect the system, as well as those objects whose attributes are changed by the behaviour of the system. 'System environment' thus depends on the system characterization, on what is comprised in the system specification and what is instead classified as external environment. With reference to Ackoff (2000), the environment of every system includes three levels of purpose: the purpose of the system; its parts; and of the system of which it is a part, the suprasystem, or highly complex system.

2.3. From systems theory to complex systems

Research underlying what is currently called 'complexity' started many years ago. Its origins, according to Prigogine (1995) lie in 'dissipative structures', or how regimes of order come into being and retain their forms amidst a constant dissipation of resources and energy. In early 1960, this idea became popularized as general systems theory (von Bertalanffy, 1956; Miller, 1978) and open systems (Kast &Rosenzweig, 1972) all of whose applications were basic and foundational.

In the same period of time (i.e., after 1960) researchers in a wide variety of subject areas were experimenting with non-linear models of dynamic systems. Several major fields of study were born of these explorations, including cybernetics (Wiener, 1948; Wiener, 1965), system dynamics (Forrester, 1958; Maruyama, 1963), computational genetic algorithms (Neumann & Burks, 1966), complex adaptive systems (Holland, 1975), deterministic chaos theory (May, 1976), catastrophe theory (Zeeman, 1977), and synergetic (Haken, 1977). More recently, Lewin and Bak (1993) and Waldrop (1993) have developed syntheses of some of these models using 'complexity' as an overarching framework.

Modern science popularized the interpretation of the simple phenomena of physics. A reductionist science par excellence, emerged as the prominent example of how the apparent chaos of the phenomena surrounding us can be make sense of by the human mind. The reason for the early success of physics was its ability to study objects that could be set out in terms of a mere few variables. The variables could be kept separated from their environment, with specifically-targeted, reproducible experiments performed on them. The dream of a 'theory of everything' drives the quest for the ultimate building blocks of the universe and for the explanation of its origin, an endeavour constituting one of the frontiers of science (Miguel *et al.*, 2012, p.2).

However, Anderson (1972) stated that a constructionist hypothesis could not by any means be implied by a reductionist hypothesis. A constructionist hypothesis will be broken down when confronted with the twin difficulties of scale and complexity. Most of the subjects or objects of scientific enquiry express these difficulties. For instance, a mere few variables cannot describe a living being. Furthermore, a human being cannot be isolated from society without changing its basic nature, and the functionality of the human body emerges from a network of interacting cells. These are examples of what are currently called ‘complex systems’. A developing body of knowledge is being collected about these complex systems, with a large number of researchers struggling to obtain a deeper knowledge of their features and the necessary sets of concepts and tools required to deal with them. As Helbing (2008) has written, these developments are gradually leading to a coherent and fundamental science of complexity.

Conventional system analysers claim that simple systems behave simply, complex systems express complex behaviour and different systems behave differently. However, this claim could not be considered universally correct. Understanding a complex system’s function, structure and response to any excitation is fundamental and the basic problem of setting the level of complexity and detail can lead to conflicting difficulties. As implied from the very name ‘complex systems’, as much data as possible should be collected to model a complex system, unless accuracy is being sacrificed for more simplicity. However, the safety and security of systems are areas where simplicity cannot take priority over complexity. This issue reflects the difficulty of finding a balance and compromise between avoiding both over complexity and over-simplification. The issue, of course, is also related to the problem of finding the right variables as mentioned previously.

The current literature about complexity of systems, defines complexity as:

- The degree to which a system's design or code is difficult to understand because of numerous components or relationships among components (ISO/IEC 2009).
- Consisting of interdependent, diverse entities that respond to their local and global environments (Page 2009).

Sillitto (2009) described the inability of a human mind to grasp the whole of a complex problem and predict the outcome as behavioural or subjective Complexity. While there are ways to reduce this complexity and improve the fit of technical systems into the complex environment, they are not the focus of this primer. Sillitto's Objective Complexity describes technical or system characteristics that lead to the subjective complexity or difficulty. As systems engineers, we have the ability to modify these characteristics; they are also the ones most frequently addressed by complex systems science.

The conventional approach to solving complexity breaks down a problem into parts, recursively, until the parts are simple enough that we understand them and can design solutions; we then re-assemble the parts to form the whole solution. The approach works well for systems whose parts interact in fixed ways (also known as "complicated" systems—an example might be a car), even when there are many interacting parts and the systems may have unpredicted behaviour (Sheard et al., 2015).

In the new system frame of data engineering science, the primary problem is often not data availability but the challenge of extracting relevant knowledge from the available big data and in devising useful data acquisition for understanding the behaviour of a system.

2.3.1. *Complex versus complicated systems*

Complicated systems have a large number of sub-systems and components which integrate in a well-understood way and have well-defined roles leading to a promised effect, e.g. modern plants with millions of physical parts, or millions of lines of software coding. Complex systems also have a large number of components, where their interactions lead to collective emergent behaviours that cannot be derived as a simple resultant from the individual components' behaviour. Prominent examples of complex systems are our brain or an airplane. All domain-based sciences such as biology, chemistry, physics, robotics, medicine and so on, study complex systems (Helbing, 2008).

Whilst there are many reasons as to why a system might be considered complex, there is no agreement on what should be the precise definition of complex. Table 2.1 presents a list of features typical of complex systems, along with some examples of systems displaying those aspects (Miguel *et al.,* 2012). Although many systems could inherit several if not all of these features, any one of them can make a system appear complex, but together they can make systems very difficult to understand and control. A key characteristic of complex systems is their ability to reconfigure themselves to create new systems with completely different properties.

Table 2.1. Reasons why systems might be considered to be complex

Many heterogeneous interacting parts	Cities, companies, climate, crowds, political parties, ecosystems.
Complicated transition laws	Markets, disease, transmission, cascading, failure rioting, professional training.
Unexpected or unpredictable emergence	Chemical systems, accidents, system breakdown spontaneous social initiatives, foot and mouth disease.
Sensitive dependence on initial conditions	Weather systems, investments, traffic jams forest fires.
Path-dependent dynamics	The evolution of the qwerty keyboard, racial conflicts, first to market.
Networked hierarchical connectives	Social networks, ecosystems, the internet voting systems, postal systems.
Interactions of autonomous agents	Road traffic, dinner parties housing markets, soccer games, crowd dynamics.
Self-organization or collective shifts	Revolutions, fashions, choirs, demonstrations, property rental markets.
Non-equilibrium dynamics	Fighter aircraft, share prices, the weather armed conflict, social networking.
Combinatorial explosion	Chess, commutations systems, data states for a program.
Adaptively to changing environments	Biological systems, manufacturing design, retail systems.
Co-evolving subsystems	Land-use, transportation, computer virus software.
Ill-defined boundaries	Genetically modified crops, nations, pollution, markets.
Multilevel dynamics	Companies, armies, aircraft, internet, transportation.

2.3.2. *Simple versus comprehensive model*

Complex systems do not need be complicated, although in reality they often are. Simple models are necessary to expose the fundamental mechanisms and provide the entry point for basic questions. The aim of simple models is to gain understanding of the so-called stylized facts of the system, i.e. to capture some essence of the real system, in other words to provide simplifications, abstractions or typical observations (with the caveat that simple models are not able to analyse all the foundations of a system). Another positive of simple models is that they easily facilitate analytical treatment and, therefore, give a useful and often deep insight into the mechanisms explaining the behaviour of the system. Complicated model are constructed and become more complicated from the extension of such simple models through gradually capturing additional details of the system. Simple models can also be very useful in proving that statements made about a system are incorrect. They therefore play a significant role in the validation of a process. On the other hand, simple models often prove only what everyone already supposed, for example, some basic system relationships. In this case, the model, though simple, captures some of the important features of the real system. This then serves as a starting point for more complicated models, with the hope of correctly capturing even more features of the real system.

Knowledge representation is a fundamental issue of complex systems. Many studies of what are today considered complex systems have traditionally relied on blind statistical analysis. The observation of these systems provides correlations of different types. Sometimes these correlations are considered to be some type of 'law of nature' that can be reproduced by ad-hoc modelling. To go beyond the knowledge provided by such correlations and be able to establish cause-effect implications is a fundamental challenge. This issue was raised, for example by Granger (1969),who investigated causal relations in econometric models and more recently by Hempel *et al.* (2011) who have worked in this direction, in the context of directed networks inference. However, knowledge representation still requires new approaches to data collection and interpretation.

2.3.3. *Modularity in technology*

Modularity is used to deal with system complexity as a general set of principles. By breaking down a complex system into smaller pieces which can then communicate with one another only through standardized interconnections within a standardized framework, one can remove what would otherwise be an unmanageable spaghetti tangle of systemic interconnections (Langlois, 2002). Modularity ideas are not new in the literature of systems architecture (Simon, 2003; Badwin& Clark, 2003). However, because of the increased complexity of modern technology, modularity is becoming ever more important today.

The world is full of complex systems. Nature provides a large number of complex organisms and ecosystems, and complex mechanical, organizational, intellectual and social systems have been constructed by humans. Simon defined a complex system as one made up of a large number of parts that interact in a non-simple way. In such systems, the whole is more than the sum of the parts, at least in the important pragmatic sense that, given the properties of the parts and the laws of their interaction, it is not a trivial matter to infer the properties of the whole (Simon, 1981, cited in Langlois, 2002). Therefore, complexity is a matter concerning both the total number of distinct parts the system comprises, and of the nature of the interconnections or interdependencies amongst those parts.

Reducing the number of elements and components in a system by grouping and clustering elements into, or by hiding elements within a smaller number of subsystems is a way to deal with complexity. Simon argued for the criterion of decomposability in modular design, which he offered both as an instruction for human designers and as a direction of the systems found ready-made in nature.

From a system's perspective, modularity can be considered as a continuous definition of the degree to which a system's elements can be decomposed and recombined, alongside the tightness of coupling between components. It is also the degree to which the rules of the system design enable the mixing and matching of elements and components (Schilling, 2000, cited in Ndou*et al.*, 2010). Modularity permits components to be produced separately, or loosely coupled (Sanchez & Mahoney, 1996) and used interchangeably in different configurations without compromising system integrity (Garud&Kumaraswamy, 1995). Modularity has also has been applied in managing complex organisations (Cusumano, 1997). For example, Mikkola (2003) used modularity to interface shared components in standardized product architecture.

In a world of change, modularity is generally worth the cost. The real issue is normally not whether to be modular, but how to be modular. Which modularization, which structure of encapsulation boundaries, will result in the best system decomposition? The aim is to find the modularization that minimizes interdependencies and most cleanly decomposes the system. The question therefore is how to do this modularization? How can we find the natural boundaries of isolation (Langlois, 2002)? The issue of defining the boundaries of encapsulation is the main challenge in the system dynamic setting. This subject needs more research and this thesis attempts to define a new approach in the encapsulations of boundaries. The research findings will present a solution that removes all the pre-specified boundaries of isolation that exist in complex systems.

2.4. Basic system definitions

In this section, some fundamental definitions in the field of systems theory, which will be used in the following chapters, will be explained. These definitions have predominantly been taken from (Cassandras & Lafortune, 2008; Ramadge & Wonham, 1989; Zeigler*et al.*, 2000).

2.4.1. *System model*

Scientists and engineers are concerned with the quantitative analysis of systems and the development of techniques to design, control, and provide explicit measurements of system performance based on well-defined criteria. Therefore, the purely qualitative definitions given in section 2.2 are not enough. Instead, a model of an actual system is sought. Intuitive thought suggests thinking of a model as a device that simply duplicates the behaviour of the system itself.

Figure 2.2 illustrates the simplest possible modelling process. A system is ‘something real’ (for example, an engine, an airplane, a plant, or a human brain), whereas a model is an ‘abstraction’ (a set of mathematical equations). Often, the model can only approximate the true behaviour of a system.

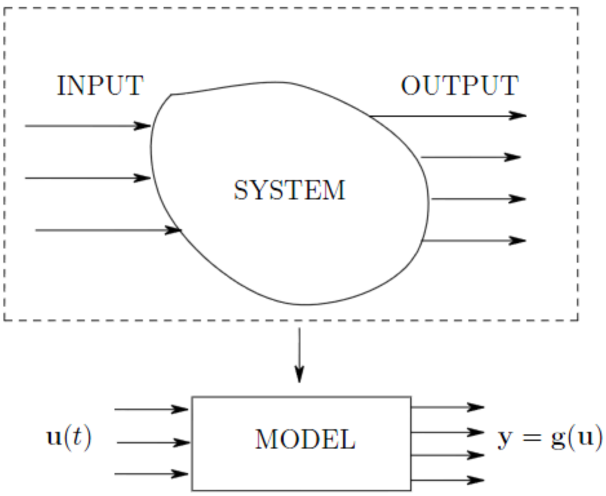


Figure 2.2. Simple modelling process

2.4.2. *Static and dynamic systems*

In static systems the output is always independent of past values of the input. In dynamic systems, the output depends upon past values of the input. Differential or difference equations are generally required to describe the behaviour of dynamic systems.

2.4.3. *Time-varying and time-invariant systems*

The behaviour of a time-invariant system does not vary with time. This property, also called stationary, implies that we can apply a specific input to a system and expect it to always respond in the same way.

2.4.4. *Linear and nonlinear systems*

A linear system satisfies the condition in equation 2.1.

$$F\,(\boldsymbol{a_1 u_1} + \boldsymbol{a_2 u_2}) = \boldsymbol{a_1} F\,(\boldsymbol{u_1}) + \boldsymbol{a_2} F\,(\boldsymbol{u_2}) \qquad \textbf{Equation 2.1}$$

Whereu_1, u_2 are two input vectors, a_1,a_2 are two real numbers, and F(u) is the function.

2.4.5. *Continuous -state and discrete-state systems*

In continuous-state systems, the state variables can generally take on any real (or complex) value. In discrete-state systems, the state variables are elements of a discrete set (for example, non-negative integers).

2.4.6. *Deterministic and stochastic systems*

A system becomes stochastic whenever one or more of its output variables are a random variable. In this case, the state of the system is described by a stochastic process, and a probabilistic framework is required to characterize the system's behaviour.

2.4.7. *Discrete-time and continuous-time systems*

In a continuous-time system, all input, state, and output variables are defined for all possible values of time. In discrete-time systems, one or more of these variables are defined at discrete points in time only, usually as the result of some sampling process (Cassandras &Lafortune, 2008, p. 46).

2.4.7.1. Discrete-Time Systems

The basic notion of time in the physical world is the assumption of time as a continuous variable. If the input and output variables of a system are defined at discrete time instants only, the result is called a discrete-time system. In contrast to the continuous-time systems considered up to this point there are several good reasons why we might need to adopt the approach of a discrete-time system (see figure 2.3). One of the main reasons is that some systems are inherently discrete-time, such as economic models based on data that is recorded only at discrete, regular intervals.

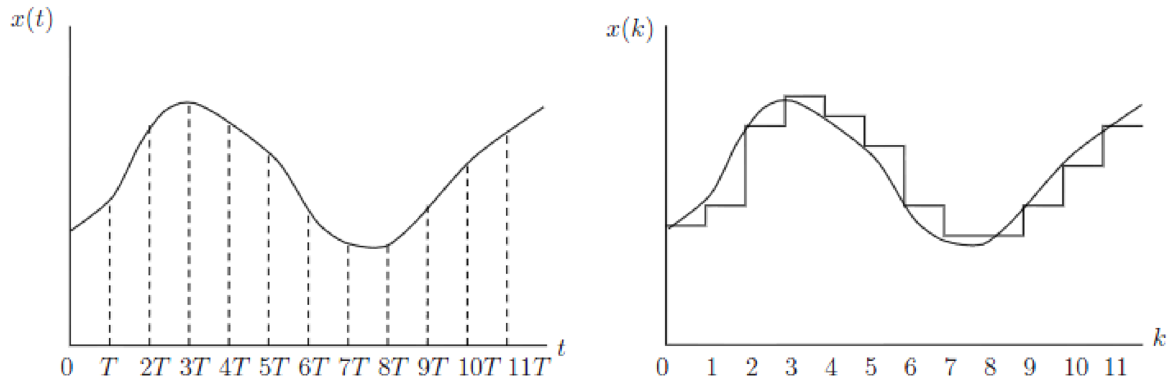


Figure 2.3. Continuous-time and discrete-time sample paths

2.4.8. *The concept of events*

An event may be identified with a specific action taken (for example, somebody presses a button). It could be a spontaneous occurrence dictated by nature (for example, a plant shuts down for reasons too complicated to comprehend), or it may be the result of several conditions which are suddenly all met.

2.4.8.1. Discrete-event systems

When the state space of a system is naturally described by a discrete set like {0, 1, 2, . . .}, and state transitions are only observed at discrete points in time, we associate these state transitions with 'events' and talk about a discrete-event system (DES). Discrete-event systems will be discussed further in the following chapters.

2.4.8.2. Time-driven and event-driven systems

In time-driven systems, the system state continuously alters as time changes. In event-driven systems, it is only the occurrence of asynchronously generated discrete events that forces instantaneous state transitions. The state remains unaffected between eventoccurrences. Figure 2.4 illustrates a comparison of sample paths for continuous systems and discrete-event systems.

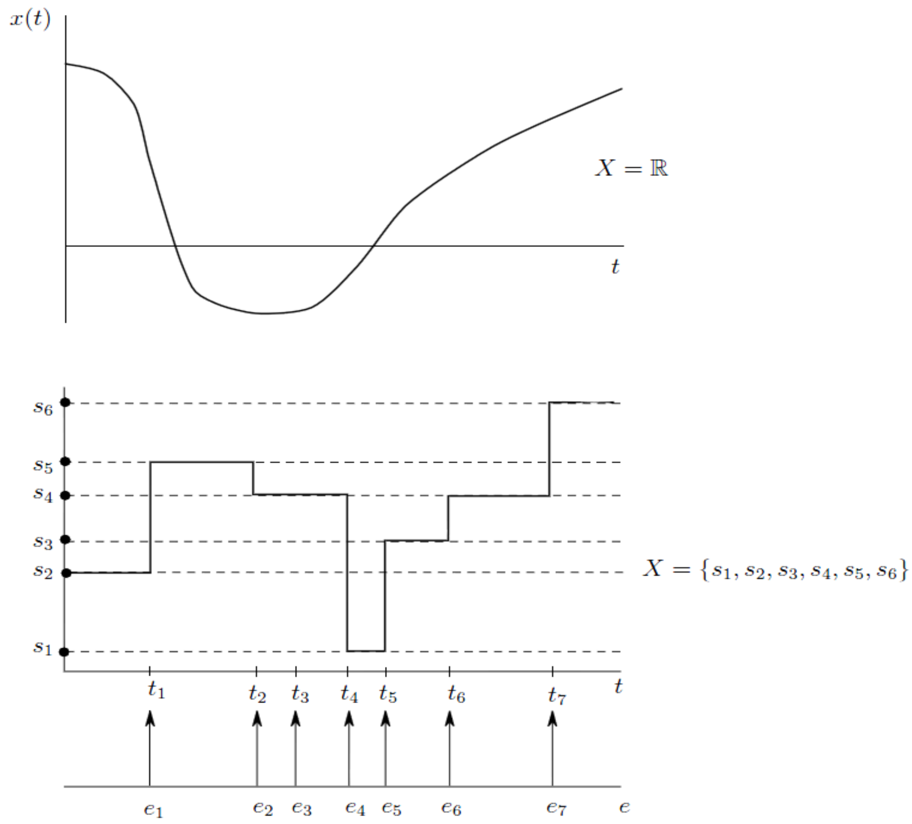


Figure 2.4. Comparison of sample paths for Continuous-Variable Dynamic Systems (CVDS) and Discrete Event Systems (DES)

2.4.9. *Response times*

The time between the presentation of a set of inputs to a system (excitation) and the realization of the required behaviour (response), including the availability of all associated outputs, is called the response time of the system.

2.4.10. *Real-time systems*

A real-time system is one whose logical correctness is based on both the correctness of the outputs and their timeliness (Laplante, 1993).

2.5. The concept of system control

The concept of system control has so far been limited to the basic issue of what happens to the system output with reference to a given input. Systems, however, do not normally work in a vacuum. In fact, it has been seen that every definition of a system contains the idea of performing a particular function. In order for such a function to be performed, the system needs to be controlled by choosing the right input so as to fulfil a desired behaviour (Dorf & Bishop, 2011).

Therefore, the input to a system is often presented as a control signal aimed at achieving a targeted behaviour (Figure 2.5). Conceptually, for a simple scalar case, this desired behaviour can be shown by a reference signal r (t), and the control input to the actual system as:

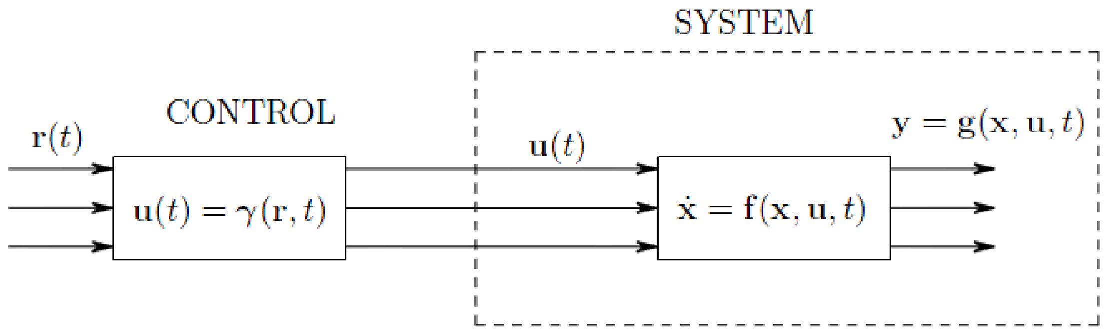


Figure 2.5. State space modelling with control input

This relationship is referred to as a control law. Thus, given the function r(t) describing a desired behaviour for the system, the task as a controller is to select u(t) = γ(r(t), t) to be the input function to the system. The extension to the vector case, where multiple reference signals are specified, leads to the control law equation 2.2.

$$U\ (t) = \gamma(r,\ t) \qquad \textbf{Equation 2.2}$$

Where γ (·) denotes the column vector whose entries are the p scalar functions u_1(t) =γ_1(r (t), t),...,u_p(t) = γ_p(r (t), t).

2.6. The concept of feedback

The idea of feedback is intuitively simple: use any available information about the system's behaviour in order to continuously adjust the control inputs. Feedback is used in everyday life in different forms. When driving, the car's position and speed are monitored by the driver so as to continuously make adjustments through control of the steering wheel, accelerator and brake pedals. In heating a house, a thermostat is used which senses the actual temperature in order to turn a furnace on or off.

2.6.1. *Open-Loop and Closed-Loop systems*

The possibility of using feedback in the controlled system model of figure 2.6 leads to one additional system classification. A system with a control input of the form u (t) = γ(r (t), t) is referred to as an open-loop system. In contrast, a system with a control input of the form u (t) = γ(r (t), x (t), t) is referred to as a closed-loop system. The distinction between these two forms of control is fundamental. In open-loop control, the input remains fixed regardless of the effect (good or bad) that it has on the observed output. In closed-loop control, the input depends on the effect it causes on the output. In the closed-loop case, it is assumed that the information feedback is some function of the state variables (not explicitly shown in the diagram), which is then included in the control law γ (r, x, t). Note that it may not be desirable to include information about all the state variables in the model. The loop formed by the feedback process in figure 2.6 gives rise to the term 'closed-loop' system.

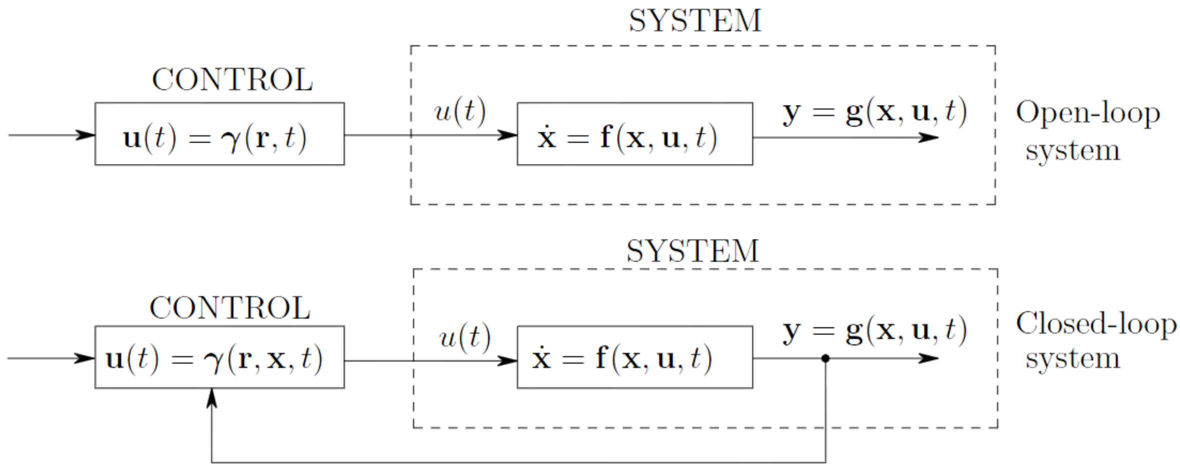


Figure 2.6. Open-loop and closed-loop systems

2.7. Summary of systems theory

This chapter has reviewed the origins and concepts of system theory, complex systems and has also addressed some fundamental definitions in systems theory and modelling.

Throughout system history, as the complexity of systems and their contexts has grown, current methods and tools have increasingly fallen short of what is needed in the face of this reality. A common approach been to seek clever ways to simplify, or reduce, the subjective complexity so that the problem and the system are understandable. Scientific advances have, in fact, often come from elegant simplifications that model the important variables or forces that dominate behaviour. However, this is not always possible – complexity often cannot be simplified away without losing the essence of the problem or possible solutions. Further, this simplification leads to an inability on the part of the solution to be able to engage with the complexity that remains despite our preference to assume it away.

This system knowledge leads this research to define and present a solution that removes all the logical boundaries of isolation that exist in complex systems. This solution comes from knowledge of systems behaviour with regard to their excitations. The following chapter reviews input selection approaches which are related to a system's inputs/stimulus section. Data integration with regard to selecting appropriate inputs to a system will lead this thesis to a novel technology which will be proposed in chapter four.

3. A Review of the Input Variable Selection (IVS)

Systems are becoming increasingly complex as they are developed to meet the requirements of possessing the necessary intelligence and physical capabilities to maintain system integrity when faced with perturbations in their operating environment that arise from both known and unknown internal and/or external event sources. The infrastructure and complex knowledge built within and around these information systems generally struggles with (a) the large flow of raw input data that clogs the communication channels; and (b) the complexity of algorithms and methods that are designed to interpret and control the state of such systems. The computational effort thus increases exponentially, leading to higher energy and time-to-action costs (Tavakoli, 2010). This effort could be reduced by identifying and filtering inputs of lesser important data.

In this chapter a general overview of complex system boundaries are presented, and the existing concepts and methodologies which could help in the reduction of the computational time of data integration and improving the quality of decisions are introduced. This chapter defines the concept of Input Variable Selection (IVS) and reviews existing methods to reveal how a data integration solution can be developed to encompass the effort of data acquisition and aggregation from the entire data source. At the same time the importance of quantifiable measures of data sources will be reviewed. At the end of this chapter the subject of gaps in data integration and management will be introduced. These gaps will lead the research to its key research question in section 1.1 which was 'How the new knowledge acquired through better sensing of systems internal dynamics and its interaction with its environment allow us to better define boundaries of the system?'. Replying to this question needs a deep knowledge of definition of input and output with respect to systems definition which in this chapter will be provided an appraisal of the literature.

An important factor that facilitates data interpretation and information modelling is an appreciation of the effect system inputs have on each output at their time of occurrence. The methods for such interpretation are generally referred to as Input Variable Selection (IVS) in engineering and as Sensitivity Analysis (SA) in mathematical literature. The purpose of IVS techniques are to maximize the quality of data acquisition and interpretation. For example, the human nervous system could collect data from its surroundings and translate this into highly-specialized knowledge such as colour, sound, temperature, and so on. The conversion of millions of data pieces to a smaller number of specific, critical knowledge enables a quick response to situations (Tavakoli, 2010). Furthermore, the selection of proper conversion techniques could improve the quality of decisions. Therefore, the cause-effect relationship between the input variables and performance parameters generates the knowledge about the system. The term ‘important’ with reference to IVS has been translated into two separate categories, those of ‘usefulness’ and ‘relevance’ (Kohavi& John, 1997; Blum & Langley, 1997).

In the following sections, brief introductions to the most relevant IVS and SA methods are provided.

3.1. IVS and feature selection

Attempts to select relevant inputs within the time-constrained system to avoid computational overhead due to redundant inputs are categorised within IVS methodologies. The objective of variable selection is three-fold: to improve the prediction performance of the predictors, to provide faster and more cost-effective predictors, and to provide a better understanding of the underlying process that generates the data (Caruna, 2003). Furthermore, there are additional potential benefits regarding variable selection: facilitating data visualization and data understanding, reducing measurement and storage requirements, reducing training and utilization times and defying the ‘curse of dimensionality’ (COD) to improve prediction performance.

Feature Selection (FS) is also classified as one of the most prevalent techniques to reduce dimensionality amongst practitioners. Although the concept of IVS and FS appear similar, there are marginal differences between them. An ‘input variable’ is generally referred to as a piece of information about the system which is later used to continuously represent the model of the system, whilst ‘feature’ refers to particularised knowledge about a series of data in the system. Therefore, ‘feature’ could be created locally or temporarily to simplify decision making systems. Local and non-sequential data could be useful in data mining (Hand et al., 2001), but a continuous flow of input data is critical in certain applications.

FS aims to choose a small subset of the relevant features from the original ones according to certain relevant evaluation criteria (Bolón-Canedo et al., 2013; Elghazel&Aussem, 2013) which usually leads to better model interpretability. A feature could be derived either from input variables (Feature Construction) or based upon data mining (Feature Extraction) (Tavakoli, Mousavi &Posland, 2013).

Feature selection has been successfully applied in a large number of applications and research, such as pattern recognition (Gheyas& Smith, 2010), text categorization (Feng*et al.*, 2012; Uğuz, 2011), image processing (Dai *et al.*, 2014), bioinformatics (Yu & Liu, 2004), biomedicine (Wang *et al.*, 2014), and so forth. For example, Tabakhani*et al.* (2014) proposed a new method in FS based on ant colony optimisation which has a low computational complexity and thus can be applied to high dimensional datasets. Their method seeks to find the optimal feature subset through several iterations without the use of any learning algorithms. Moreover, the feature relevance has been computed based on the similarity between features, which leads to the minimization of redundancy.

IVS can therefore be described as a direct result of the collection of raw input data from the data source. In distinction, FS is knowledge extracted by the mining of data collected from input variables. Regression (Uysal&Güvenir, 1999) and cluster analysis (Jain, 2010) are two of the most used techniques in dimensionality reduction. The advantages and disadvantages for knowledge discovery using regression techniques are discussed in the following sections, whilst cluster analysis techniques will be explained in the next chapter.

3.1.1. *Regression*

Hand *et al.* (2001) defined regression as the task of estimating relationships between a dependent variable and a number of independent variables. Independent variables could be assumed as input variables of a system, whilst, the dependent variable is interpreted as the model's key performance indicator (KPI). Jain (2010) used the regression method for predicting and learning the numeric features of a model when there is no predictive model between independent and dependent variables. This method has also been used to build a new set of independent variables to replace the original set of independent variables which leads to the same effect. The term 'derived variable' has been created for this purpose by the data mining community. Well known methods within data mining techniques in this area are Projection Pursuit Regression (PPR) and Principal Component Analysis (PCA). PPR forms an estimation model by transforming the training set onto lower dimensional projections as a solution for high dimensional data sets and PCA is a multivariate orthogonal transformation cluster analysis (Hassan &Habeb, 2012).

The regression method can be seen to be invalidated by the heterogeneous nature of data distribution. It also becomes quickly and extremely unreliable in high dimensional input data analysis; this phenomenon is called the 'curse of dimensionality' (COD). Banks *et al.* (2003) describe how the number of possible regression structures increases faster than exponentially with dimensionality. In comparing the performance of regression methods by conducting experiments on ten well known regression methods the results of their research concluded that no particular regression method is capable of supporting the scale and heterogeneity of variables in volatile industrial systems whilst keeping the computational cost low.

In the context of IVS, the relationship between the input variables and system performance are considered as system knowledge. In this context, input variables with different levels of relationships are classified into different categories. From another viewpoint, some functions are required to measure and study the level of impact of each independent variable on the dependent variable or system performance. In the following section, a discussion with respect to sensitivity analysis will be opened up.

3.2. Sensitivity analysis (SA)

In simulation models, assumptions represent uncertain information regarding the model that cannot be obtained from the system in reality. These model assumptions could be about a variable value or a heuristic decision by the system manager. The effect of input uncertainty on the model output is evaluated by Sensitivity Analysis (SA) (Saltelli*et al.*, 1999). SA is a technique to minimize the cost of data acquisition and subsequently its interpretation, by eliminating the input variables that have the least impact on the system (Volkova, Iooss & Van Dorpe, 2008; Cloke, Pappenberger & Renaud, 2008; Hu & Shi, 2010; Tavakoli, Mousavi&Broomhead, 2013; Fock, 2014). The purpose of sensitivity analysis methods is to measure the true impact of a system's input on a system's output. However, focusing only on the most valuable information has a significant impact on the behaviour of systems. Sensitivity indexing is a systematic approach for expressing the relationships between the inputs and outputs of a system. However, due to the epistemic uncertainties of system input-output relationships finding a true representation is challenging (Krzykacz-Hausmann, 2006).

The selection of a suitable sensitivity analysis method according to (Tavakoli, Mousavi&Broomhead, 2013) requires firstly to be aware of the relationships between input and output variables. The selection of appropriate methods for sensitivity analysis are classified by three major factors. These factors are: the analytical relationship between input and output data; the statistical distribution of input variables; and finally, the computational overhead.

3.2.1. *The analytical relationship between input and output data*

Analytical SA methods attempt to describe the impact of changes in one variable in relation to others using analytical models. SA methods such as differential analysis, coupled/decoupled direct and Green's function are categorised as various analytical SA methods (Saltelli, 2002). In those cases where mathematical equations do not exist between the respective system variables, statistical techniques attempt to extract the relationship from input-output variables. Morris (Jin*et al.*, 2007), the Fourier Amplitude Sensitivity Test (FAST) (McRae *et al.*, 1982), Monte Carlo (Sobol, 2001) and Latin Hypercube (Hora& Helton, 2003) are some examples of these techniques. Their reliance on historical data and the generation of data samples which fit to probabilistic equations are prominent characteristics of these techniques. The entropy method (Krzykacz-Hausmann, 2006) is a method that is less reliant on analytical methods for extracting sensitivity indices. In the following sections these methods are explained.

3.2.2. *The statistical distribution of input variables*

Input data series distribution normally influences the sensitivity indices of a system. For example, some sensitivity analysis methods, such as those based on linear regression, can inaccurately measure sensitivity when the system performance is nonlinear with respect to its inputs. Correlation-based SA methods (Cohen*et al.*, 2013) are not able to recognise nonlinear relationships between the input and output series of a model. In such cases, variance-based measures are more appropriate. Variance-based and entropy-based indices are sensitive to heteroscedastic data series (Krzykacz-Hausmann, 2001).

Saltelli and Annoni (2010) classified two more constraints in the choice of SA methods:

- Input correlation: correlation between input variables must be considered. Most sensitivity analysis methods assume the model inputs are independent.
- Parameter interactions: the excitation of two inputs at the same time causes variation in the output larger than that of varying each of the inputs alone, causing interactions to happen. Such interactions are present in any model that is non-additive, but will be neglected by methods such as one-at-a-time (OAT) perturbations. A total order sensitivity index can measure the effect of such interactions.

3.2.3. *The computational overhead*

SA methods are hungry to computational processes. The majority of SA methods capture a large batch of input variables running a number of times. Subsequently, SA values are measured based on the number of iterations needed to perform the SA algorithm. Whilst different methods have their own approach to determine the SA indices, their commonality is a requirement for extensive computational processing.

Tavakoli, Mousavi and Broomhead (2013) proposed the EventTracker sensitivity analysis which has a clear advantage over analytical and computational IVS methods since it tries to understand and interpret system state changes in the shortest possible time with minimum computational overhead. The objective of this thesis is also proposes a novel sensitivity analysis method for time-constrained applications.

Related works

Figure 3.1 presents the most widely used SA methods in three different categories regarding their input/output relationships. In this section a review of the latest analytical, sample-based and heuristic SA techniques is provided.

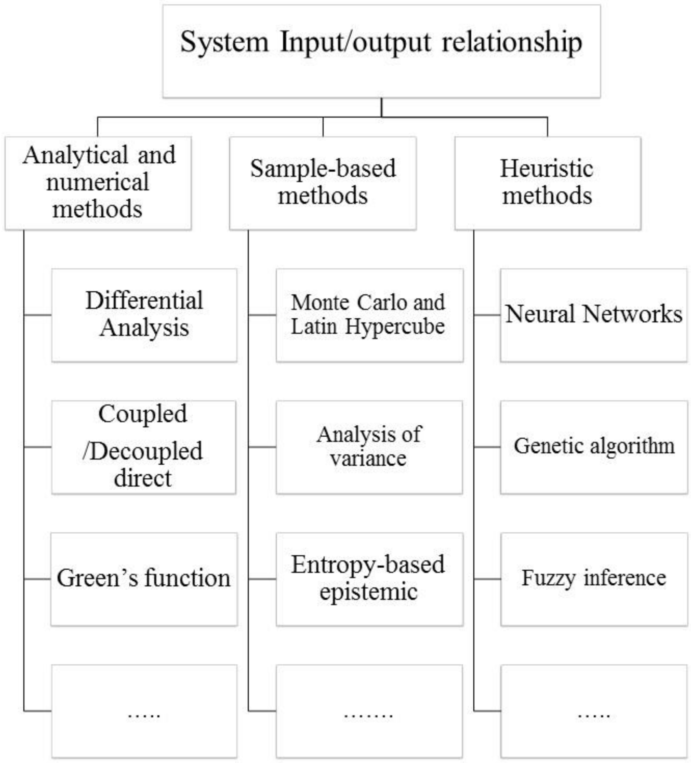


Figure 3.1. Different SA methodologies

3.2.4. Analytical methods

Analytical and numerical methods measure the impact of change in one variable on the others by means of a mathematical equation that describes the relationship between variables. Isukapalli (1999) classified differential analysis with Green's function and couple/decoupled direct methods both widely-used analytical methods.

3.2.4.1. Differential analysis

Differential analysis, also referred to as the direct method, is explained here first because this method provides the foundation for almost all the other sensitivity analysis techniques.

A sensitivity coefficient is essentially the ratio of the change in output compared to the change in input whilst all other parameters remain constant (Krieger *et al.*, 1977). This scenario is defined as the 'base case' of a model. Differential techniques are structured on the behaviour of a model for a base-case scenario, i.e., all parameters are set equal to their mean value. Differential sensitivity analysis is based on partial differentiation of the aggregated model. When an explicit differential equation describes the modelled relationship, the sensitivity coefficient for a particular independent variable is calculated from the partial derivative of the dependent variable with respect to the independent variable (Caruana & De Sa, 2003). Methods such as the Neumann expansion and perturbation methods (Buonomo & Lo Schiavo, 2010) could help to extract these coefficients by approximating differential equations. However, complex and nonlinear relationships between system variables cannot be guaranteed in this type of analysis.

Morisawa and Inoue (1974) used the differential method to select desirable conditions for underground waste disposal sites in Japan. However, they reported that with the direct method, the magnitude of variable sensitivity is dependent upon the base-case scenario. A major drawback is that this localized behaviour may not be applicable for areas far removed from the base case.

3.2.4.2. Green's function

In the Green's function method, the sensitivity equations of a model are obtained by differentiating the model equations. The sensitivity equations are then solved by constructing an auxiliary set of Green's functions. This method minimizes the number of differential equations that are solved for sensitivity, and replaces them with integrals that can be easily calculated (Isukapalli, 1999). The concept of Green's function stems from the knowledge that the total output of a linear time invariant system can be formulated by a summation of terms that adds all outputs of the system for all single points (Beylkin*et al.*, 2008). Disadvantages regarding the application of Green's function are its constraint to a linear and time-invariant system, and also its ability to work only with ordinary differential equations which govern dependent variables with respect to independent variables. In real applications it is often difficult to separate independent variables and dependent variables. Additionally, working one variable at a time for multi-dimensional systems could be computationally expensive (Tavakoli, Mousavi & Poslad,2013).

3.2.4.3. Coupled/Decoupled direct method

The coupled direct method involves the differentiation of model equations and the subsequent solution of the sensitivity equations. The sensitivity equations are then solved along with the original model equations (Coupled Direct Method) or separately (Decoupled Direct Method). The decoupled method is reported to be more efficient than the Green's function method (Isukapalli, 1999). In common with other analytical methods, prior knowledge of the model equations is a requirement. The couple/decoupled methods also exhibit the feature of being model-oriented and expert-hungry. These features make them less attractive for practical applications when compared to SA methods that do not require model equations.

3.2.5. *Sampling based methods*

Sampling based methods do not require access to model equations. These methods run a number of models at a given set of sample points, and attempt to establish a relationship between inputs and outputs using the model results at the predefined sample points. This may be required due to the non-existence of analytical relationship between model variables, the lack of expertise to identify such a relationship, or due to changes in the configuration of inputs and outputs. Consequently, the effort required by expert interference is often costly and may vary. On these occasions, sampling based SA methods tend to establish a model equation. They do so by identifying certain statistical features in the distribution of the data series of the two variables.

The general shortcoming of sampling based methods is their reliance on historical data. Their reliability decreases when there is little time for collection and interpretation of the historical data. Cloke *et al.* (2008) applied their model to 1280 sample values of 20 input parameters. Each cycle of sample generation and model execution took between 2 and 52 h per set of samples. The overall execution cycles took almost 46 days. This example reveals the significant shortcoming of sampling based analysis for volatile systems that require quick analysis and reaction.

Some of the most widely used sampling based sensitivity/uncertainty analysis methods are: Monte Carlo and Latin Hypercube Sampling, Analysis of Variance (ANOVA), Fourier Amplitude Sensitivity Test (FAST), Sobol and Entropy-based Epistemic sensitivity analysis.

3.2.5.1. Monte Carlo and Latin Hypercube methods

The Monte Carlo method is one of the most widely used techniques for uncertainty analysis. This method performs sampling from a possible range of input variable values followed by model evaluations for the sampled values until a statistically significant distribution of outputs is obtained. An important part of every Monte Carlo analysis is the generation of random samples. These generating methods produce samples from a specified distribution (typically a uniform distribution). The random numbers from this distribution are then used to transform model parameters according to a predetermined transformation equation (Griensven *et al.*, 2006).

Problems such as optimization and simulation can be addressed through the Monte Carlo analysis method. Since this method requires a large number of samples and/or model runs, their applicability is sometimes limited to simple models. In the case of computationally intensive models, the time and resources required by this method could be prohibitively expensive. To mitigate such an overhead a degree of computational efficiency is accomplished by the use of the Modified Monte Carlo (MMC) method which samples from the input distribution in an efficient manner (Andrieu *et al.*, 2010; Liu *et al., 2013).*

The Latin Hypercube Sampling (LHS) method was first developed by (McKay & Conover, 1979) and is one such widely used variant of the standard Monte Carlo method. In this method, the range of probable values for each uncertain input parameter is divided into intervals of equal probability. Thus, the whole parameter space, consisting of all the uncertain parameters, is partitioned into cells having equal probability, and these are sampled in an 'efficient' manner such that each parameter is sampled once from each of its possible intervals. The procedure of LHS for selecting K different values from each of N variablesX_1,X_2, ..., X_N can be summarised as: (i) Divide the range of each variable into K at equal intervals; (ii) From each interval, randomly select a value with respect to the probability density in the interval; (iii) The M values thus obtained for X_1 are paired randomly with the K values ofX_2. These K pairs are combined in a random manner with the K values of X_3 to form K triplets, and so on, until a (K × N) matrix is formed. The advantage of this approach is that the random samples are generated from all the ranges of possible values, thus giving insight into the extremes of the probability distributions of the outputs(Hora & Helton, 2003).

An important challenge faced when applying Monte Carlo methods in time-sensitive applications is the effort required to estimate the distribution of the input variables prior to sample generation. This can be computationally very expensive, particularly with high dimension input variables. LHS sampling can generate the input distribution with fewer sampling iterations than the Monte Carlo sampling to achieve a similar accuracy (Zi, 2011).

3.2.5.2. Analysis of variance (ANOVA) methods

The One-At-a-Time (OAT) method of processing input variables can at times make it impossible to capture the complexity of the relationship and interaction that exists between multiple input variables and an output variable. The ANOVA method aims at decomposing and measuring the variance of the output distribution, when all inputs are varying, into partial variances (Saltelli*et al.*; 2000; Armstrong *et al.*, 2000; Volkova *et al.*, 2008).

For the output of a system represented as an analytic function of input variables, e.g. Y=$f(X_1, X_2, \dots, X_p)$, the relative importance of the independent inputs can be quantified by the fractional variance which is defined as the fractional contribution to the output variance due to the uncertainties in inputs. This can be estimated using an ANOVA decomposition formula for the total output variance Var (Y) (Yu and et al., 2009).

$$V = \mathrm{Var}(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \cdots + V_{12\cdots p}$$

Where **Equation 3.1**

$$V_i = \mathrm{Var}(E(Y|X_i = x_i)) \text{ and}$$

$$\begin{aligned} V_{ij} =& \mathrm{Var}\left(E(Y|X_i = x_i, X_j = x_j)\right) \\ & - \mathrm{Var}(E(Y|X_i = x_i)) - \mathrm{Var}(E(Y|X_j = x_j)) \end{aligned}$$

Equation 3.2

A partial variance V_i represents the main, or first order, effect of an input i on the output that corresponds to the variance when all other inputs are constant. Higher order effects $V_{1,2,\dots,p}$are combined effect for 2 or more inputs. The partial effects can be estimated with special sampling schemes that are often computationally hungry.

Where E($Y/X_i = x_i$) denotes the expectation of conditions on X_i having a fixed value X_i, and V stands for variance over all the possible values of X_i.Sensitivity index of output to each input variable is (Saltelli, 2002):

$$\boldsymbol{S_i} = \frac{\boldsymbol{V_i}}{\boldsymbol{V(y)}}$$ **Equation 3.3**

To achieve these variables, when no explicit relationship exists between inputs and output, a numerical approach based on sample generation can be adopted. Using this technique, the level of computational overhead, in terms of model runs required to produce output values for each input sample grows rapidly. The amount of computational overhead, in terms of the number of model runs (for producing output values per each input sample set) can be derived using the equation 3.4.

$$M = N \times \sum_{i=0}^{p} \frac{p!}{(p-1)!i!}$$ **Equation 3.4**

Where N is the sample size, and p is the number of input variables (Saltelli, 2002),for instance, with 100 samples and 5 input variables, the number of execution runs is 1,358, which is extremely high. Therefore this method would not be applicable in real-time applications.

The computational cost of a numerical calculation is defined in terms of the number of model runs necessary to estimate the sensitivity measure. The cost for computing all terms in the variance decomposition is given by (Rabitz&Aliş, 1999).

3.2.5.3. Fourier Amplitude Sensitivity Test (FAST) Method

The Fourier Amplitude Sensitivity Test (FAST) method was proposed by Cukier *et al.* (1973) to study chemical reaction systems. The FAST is a variance-base method based on Fourier transformation of uncertain model parameters into a frequency domain, thus reducing the multidimensional model into a single dimensional one. FAST is an example of improvements in computational efficiency of the ANOVA-based SA methods. In opposition to ANOVA methods, the data distribution of input variables cannot be estimated from the acquired historical data. Instead, all distributions of input variables are considered to be uniform and within a specific range. Therefore, generated samples in this range follow a periodical function (Minnebo *et al*, 2007).

The FAST method assumes that all model parameters are independent from each other. The parameter is sampled from the following transformation function (Zi, 2011).

$$\boldsymbol{P_i = P_i^o e^{u_i} = P_i^o e^{G_i(\sin w_i s)}}$$ **Equation 3.5**

Where$p_i{}^0$is the reference value for parameter i.

S is a scalar variable.

w_i is an element in a set of linearly independent integer frequencies.

Gi is a defined transformation function that transforms the probability density of the parameter into s space.

If assumes $f(s)$=$f(G_1(\sin w_1 s), G_2(\sin w_2 s), \ldots, G_k(\sin w_k s))$.

The model output can be calculated by

$$\widehat{\boldsymbol{E}}(\mathrm{y})=\frac{\mathbf{1}}{\mathbf{2}\boldsymbol{\pi}}\int_{-\boldsymbol{\pi}}^{\boldsymbol{\pi}}\boldsymbol{f}(\boldsymbol{s})\mathbf{ds}$$ **Equation 3.6**

The partial variance in FAST method is approximated by:

$$D_i^{\mathrm{FAST}} \simeq 2\sum_{p=1}^{\infty}(A_{pw_i}^2 + B_{pw_i}^2)$$ **Equation 3.7**

$$A_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)\cos(js)\,\mathrm{d}s \qquad B_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(s)\sin(js)\,\mathrm{d}s$$

The FAST sensitivity index is calculated with the following approximation

$$S_i^{\mathrm{FAST}} = \frac{D_i}{D} \simeq \frac{2\Sigma_{p=1}^{\infty}(A_{pw_i}^2 + B_{pw_i}^2)}{2\Sigma_{j=1}^{\infty}(A_j^2 + B_j^2)}$$ **Equation 3.8**

Saltelliet *al*. (1999) developed extended FAST (eFAST) which allows the computation of the total contribution of each input factor to the output's variance. The eFAST method has a better transformation function than the classic FAST method because it provides uniform distributed samples for the parameters. The main advantage of eFAST is its robustness, especially at low sample size and its computational efficiency. However, Marino *et al.* (2008) implemented eFAST in a biological case and indicated its computational cost as its major drawback, especially for computing sensitivity indices. This experiment highlights how a large number of iterations are required to achieve an acceptable degree of accuracy in eFAST.

3.2.5.4. Sobol Method

In 1990 a Russian mathematician Ilya M. Sobol (1990) developed a method for SA which was considered a natural extension of the FAST approach. His method is categorized in variance-based sensitivity analysis methods because this method computes the ANOVA decomposition of the output variance.

Consider the function f (x) $=f(x_1, x_2, \dots, x_n)$ defined in the n-dimensional unit cube. Under Sobol's assumptions this function could be decomposed into summands of increasing dimensions (Saltelli & Bolado, 1997):

$$f(x_1, \dots, x_n) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum \sum_{1 \le i \le j \le n} f_{ij}(x_i, x_j) + \cdots + f_{12\dots n}(x_1, x_2, \dots, x_n),$$

Equation 3.9

Wheref_0 is a constant and the integrals of every summands over any of its own variables in zero:

$$\int_0^1 f_{i_1\dots i_s}(x_{i_1}, \dots, x_{i_s})\, dx_{i_k} = 0, \quad 1 \le k \le s.$$

Equation 3.10

The total variance of (x) can be written as

$$D = \int_{K^n} f^2(\bar{x})\, d\bar{x} - f_0^2$$

While

$$D_{i_1\dots i_s} = \int_0^1 \cdots \int_0^1 f^2_{i_1\dots i_s}(x_{i_1}, \dots, x_{i_s})\, dx_{i_1} \cdots dx_{i_s}$$

Equation 3.11

is a contribution to the total variance due to a generic term $f_{i_1 \dots i_s}$ in the series development.

Sensitivity $S_{i_1 \dots i_s}$ can be introduced:

$S_{i_1 \dots i_s} = D_{i_1 \dots i_s} / D$

$$D=\sum_{i=1}^{n} D_i + \sum \sum_{1 \le i \le j \le n} D_{ij} + \cdots + D_{12\dots n}$$

Equation 3.12

It follows that $\sum S_{i_1 \ldots i_s}$=1 which ,for example, S_1 is the main effect of parameter x_1, , S_{12} is the interaction effect, i.e. that part of the output variation due to x_1 , x_2 parameters. Finally $S_{123\,..n}$ is that fraction of the output variance which cannot be explained by summing terms of lower order. This ANOVA like decomposition is similar to the FAST one.

The advantage of Sobol with respect to FAST is its capability to compute higher-order terms in the variance series development. It is therefore relevant when those terms make a significant contribution to the output variance. However, Sobol indices are computationally more expensive, although they converge to the analytical values. Furthermore, Sobol indices provide a unique way to estimate the effect of variables as well as interaction terms of any order.

3.2.5.5. Entropy-based sensitivity analysis

Entropy is a well-known function in the theory of information, which presents the loss of information within a system and then, by way of contrast, the amount of information (Auder, 2008). The entropy of a discrete random variable x ranging inx_1,... , x_n with respective probabilitiesp_1,... , $p_n$is:

$$\mathbf{H(x)} = -\sum_{k=1}^{n} p_k \ln(p_k) \qquad \textbf{Equation 3.13}$$

This quantity does not only depend on the values of x, but also on their probabilities. Using entropy definitions, Krzykacz-Hausmann (2001) introduced the entropy–based SA method to deal with the high computational effort and time-demanding issues of sampling-based SA methods like the Monte Carlo sampling method. This method proposes an approximation approach that measures the entropy of variable distributions from original samples. It has been proposed that in order to determine sensitivity indices one needs only to establish the value of independent input variables (denoted by X) and dependent output variables (denoted by Y). The sensitivity indices are defined as:

$$\acute{\eta}_i = 1 - \frac{H(Y/X_i)}{H(Y)} \qquad \textbf{Equation 3.14}$$

Where H(Y) is the entropy values and $H(^{Y}/_{X_i})$ are the values of conditional entropy, which is a representation of the information learnt on Y by the knowledge of X.

This method replaces the time consuming sample generation of X and evaluation of Y by simple random sampling using piecewise uniform density function estimations. The entropy-based method has been applied in many areas, for example, Chen *et al.*(2014) introduced a concept called neighbourhood entropy which was established to evaluate the uncertainty of a neighbourhood information system. Consequently, the entropy-based roughness and approximation roughness measures of the neighbourhood system were presented. The results show that the entropy-based approximation roughness can provide more information for evaluating the uncertainty in neighbourhood decision systems.

3.2.6. *Heuristic based methods*

The input variable selection process based on heuristic methods normally relies on the knowledge of system experts. This knowledge manifests itself in the form of experience, engineering and modelling expertise, or special algorithms. For example, but not exclusively, methods such as Fuzzy Inference Models, Genetic Algorithms (GA) and Artificial Intelligence (AI) (e.g. artificial neural networks) all fall into this category.

The strength of heuristic methods in solving complex data modelling and control systems is well-recorded in literature and industry. For example, cement factories worldwide are being controlled by the direct knowledge of expert kiln operators. Fuzzy control of cement kilns has been one of the first successful applications of fuzzy control in industry. Expert knowledge has a direct impact on identifying the fuzzy inference rules that optimise the key performance indicators of the manufacturing process. To reduce reliance on direct expert input which is extremely time consuming and prone to variation, automatic learning methods such as AI techniques have been used. The AI techniques look into the pattern of acquired data and derive the necessary knowledge for measurement or optimization plans. GA techniques are also considered as one of the methods to derive knowledge from a known set of data points (genomes) and use the principles of random mutation and filtering of unwanted genes. The GA can be built with arbitrary flexibility and can be successfully trained using any combination of input variables (May *et al.*, 2011). For example, Madhanagopal et al. (2014) have used GA effectively for automating IVS processes in decision making units. However, this technique is reliant on a system's expert knowledge (to predefined rules) and is computationally-hungry, which is a significant shortcoming of this technique.

3.3. Conclusion

This chapter reviews the existing literature on Input Variable Selection and feature selection methods. Input variable selection and feature selection are solutions to the problem of dimensionality reductions. However, feature selection is used for more specific purposes than input variables. Replying to the key research question about removing system boundaries and meeting the research objectives, finding an IVS and FS method that does not generate extra computational effort to construct new variables, it is necessary to know the relationship between input variables and the system output (i.e. the system's performance parameters).

The method for measuring input variables based on their influence on system outputs, the so-called sensitivity analysis, can be selected based on preferences and priorities on the relationship between input and output variables, the data distribution of variables, and perhaps most importantly, the computational cost of the method. Based on these attributes, according to the knowledge of authors, with the exception of (Tavakoli, Mousavi & Broomhhead, 2013) there is no sensitivity analysis method that competently works with complex systems with regards to heterogeneity and large number of input variables in real-time (with time constraints).Therefore, a new approach in real-time sensitivity analysis methodology will be introduced in chapter four. Its various applications in different industries and comparison of its efficiency to event-based EventTracker sensitivity analysis will be discussed in later chapters.

4. Event Clustering Data Grouping Technique (EventiC)

Complex man-made electro-mechanical devices require high-quality information on which to base timely responses to events occurring in their volatile environments. These devices are required to meet the ever increasing demand on performance, responsiveness, adaptability, regain composure in cases of internal/external destabilization, and importantly work at optimal levels of energy efficiency and utilization (Bolton 2003). In other words the main challenges of design are to assemble systems that operate safely and economically regardless of the internal and external destabilizing factors. Due to the interrelatedness of modern complex systems and the demands from these systems to operate in extreme conditions, designing ever more complex and more adaptive systems is a major challenge for system designers.

In pursuit of meeting the challenges posed by the modern and evolving complex systems, recently, conceptualized and coined as Cyber-Physical Systems (CPS) technology – the Event Clustering Sensitivity Analysis called EventiC is suggested as a unique and novel data and knowledge engineering platform to meet the challenges of "*the dynamic, autonomous, adaptive and self-organizing embedded systems, and seamless and secure interaction of the embedded system/cyber-physical systems with their environment*". The proposed concept endeavours to create a logical and simple basis to manage the interrelationships and internal dynamics of the components within the eco-system of embedded systems and their environment (Danishvar *et al.*, 2014). Its sole purpose is to take the first step in understanding the causal relationships between the system and its operational environment as the system changes state and boundaries. As the first of its kind, EventiC will be able to evaluate in real-time the impact of every relevant event on the performance, control, stability and overall behaviour upon the system.

The challenge that EventiC deals with is to overcome the shortcoming of EventTracker (Tavakoli et al. 2010). Whilst EventTracker deals with 1 to many correlations, the EventiC cluster is intended to deal with many to many relationships, thus developing the most effective grouping techniques become important.

What distinguishes the proposed event clustering technique from other automated data pattern and knowledge derivation techniques is its simplicity and speed in extracting all the available data in the system domain, converting and then processing the necessary information in near real-time. There is no reliance on a set of predefined rules such as good/bad data, historical trends, or investigations into long-term patterns. More importantly, unlike Heuristic techniques, EventiC does not rely on prejudgment of data relevancy that normally emanates from expert interference – it is an unbiased method. It achieves a correlation analysis throughout the analysis span and deals with too many relationships of input and output parameters. To the best knowledge of the author, no such challenge has been achieved to date.

4.1. The basic concept of Event Clustering

The basic assumption of the Event Clustering technique (Bolton, 2003; Danishvar *et al.*, 2013; Danishvar *et al.*, 2014) is that the state of a system during its life span can be broken down into a series of consecutive discrete events. The change in the state of a system can be triggered by various events. These events are instigated by changes in the state of the input variables (sensors and actuators). In real-time this can help to group such important events with the performance indicators of the system. It is important to understand that the discrete unaware event described here implies that the system is not aware of the previous event.

The challenge of explaining or interpreting the state (being) of a physical entity (i.e. system) has fascinated philosophers, system theorists and engineers. The underpinning philosophy of the proposed technology is based on René Descartes' philosophy of 'Discours de la méthode', or "breakdown every problem into as many separate elements as possible" (Hoff, 2013), and then reassemble it in the form of an eco-system of causality of the smallest units. The concept of 'coincidentia oppositorum' or the 'fight among parts' from the renowned fifteenth-century thinker Nicholas of Cusa (Minnebo & Stijven, 2011; Hoff , 2012) has also been adopted and can be interpreted as the concept of the causal interrelationships (mathematical law) of parts in the whole.

In the language of engineering, how the excitation of the system, driven by events and demonstrated by measurable inputs that contribute/affect the behaviour, stability and safety of a system is explained and simplified in figure 4.1.

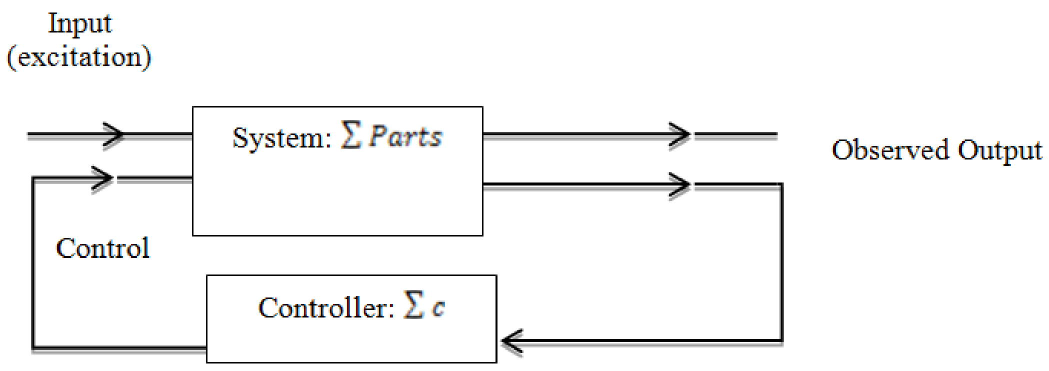


Figure 4.1. A system and controller in closed-loop

The scientific and technological challenge is to be able to assemble a system (process) definition i.e. 'Character Equation' (CE) that genuinely/accurately represents the complex system in timely fashion. The current theoretical systems and control approach has successfully followed the 'Discours de la méthode' of Descartes, and have isolated individual systems with near perfect character functions. In the realm of physical (phenomenological) models (i.e. Newtonian, Kirchhoff's circuit laws, etc.), numerical and analytical models (finite element analysis, statistical inference models, etc.), and finally where processes become difficult to explain via character functions, heuristic models (neural networks, genetic algorithms, fuzzy inference, etc.) significant achievements can be observed. However, the principle of finding a perfect solution is the assumption regarding the knowledge of all excitation parameters at the outset. Therefore control systems even for complex vehicles/processes normally become a multitude of isolated problems working independently. The challenge becomes even more complex if the system needs to understand excitation and respond to it accordingly. Generally, the three aforementioned methods come to the aid, since they rely on historical events and records. However, currently there are very few solutions that can make sense of the data in real-time and respond to the excitation in an optimal manner.

The problem of timeliness has been solved with the inclusion of controllers (e.g. micro-controllers, programmable logic controllers, etc.) within the circuitry of systems, thus creating a level of integration. At present this integration is in primitive form manifesting itself in Supervisory Control and Data Acquisition Systems (SCADA). Whilst they are powerful recorders, managers and bundlers of data they possess little added value beyond that. In effect these real-time systems, at all levels, conduct very little raw data processing and interlinking. The interlinking and analysis of data is the responsibility of higher level systems (e.g. digital controllers, neural networks, fuzzy controllers, artificial intelligence techniques, or simply direct human intervention). By the time these systems properly learn the patterns, the complex system has moved on. Only if the incident happens again, will the models be able to decipher and find the appropriate response to the excitation. Borrowing from von Bertalanff, Problems must be intuitively seen and recognised before they can be formalised mathematically" (von Bertalanffy, 1956), otherwise, mathematical formalism may impede rather than expedite the exploration of this very real problem. It is this thesis' belief that such a visualisation of the observable world is the key to solving complex problems, and it is proposed that EventiC can take a logical step towards such visualisation.

4.2. Data clustering methods for big data

Clustering is an unsupervised learning class of methods in which objects are grouped into a set of disjointed classes, called clusters, so the objects within such classes possess close similarity. The goal of data clustering, also known as cluster analysis, is to find the natural groupings of a set of patterns or events. Webmaster (Merriam-Webmaster online Dictionary, 2015) defines cluster analysis as a statistical classification technique for discovering whether the individuals of a population fall into different groups by making quantitative comparisons of multiple characteristics.

There is a vast body of knowledge in the area of clustering and large numbers of clustering algorithms have presented themselves to analyse the massive volume of data generated by modern applications. However, a review of the literature on clustering techniques reveals that despite the massive volume of algorithms used on a variety of applications such as machine learning (Bishop, 2006), k-means (Pena et al., 1999), data mining and pattern recognition (Aggrawal&Zhai, 2012), self-organising maps (SOMs) (Xiao et al., 2003),hierarchical clustering (Eisen et al., 1998), evolutionary algorithms (Hruschka, 2009), bio-informatics (Yeung, 2001) and others, it is not easy to decide the most appropriate algorithm for any given data set to satisfy both the requirements of the computation efficiency and result quality.

It is generally observed that different clustering results are produced when clustering is applied to the same dataset while adopting different clustering methods, different sets of parameters for the same method, or the same stochastic method over multiple runs (Vega-Pons & Ruiz-Shulcloper, 2011). However, there is no one superior method which overcomes all other methods in quality in all cases. Therefore, it is a common question to ask: which of those different sets of results should be considered, or initially, which clustering method should be adopted?

4.2.1. *Clustering algorithm categories*

Fahad *et al.* (2014) have conducted a comprehensive survey of the most utilised clustering algorithms on big data. Figure 4.2 shows an overview of framework of clustering taxonomy.

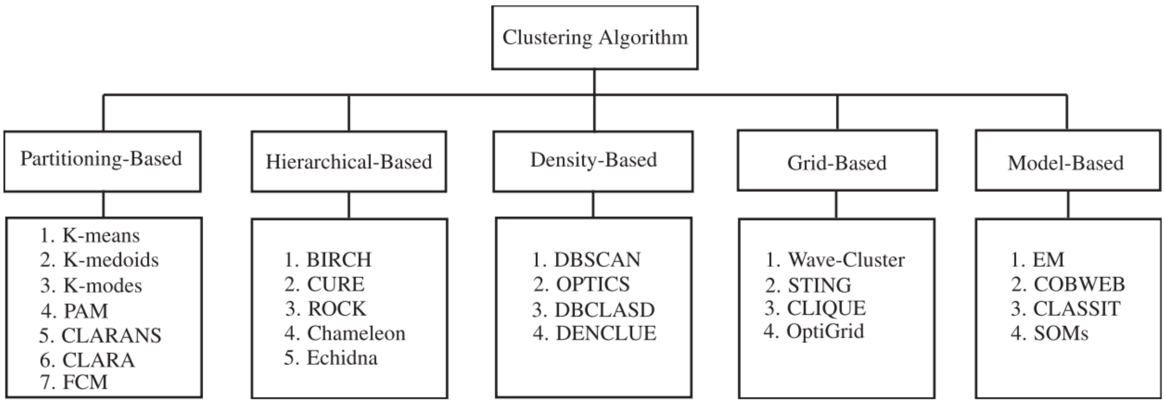


Figure 4.2. An overview of clustering taxonomy (Fahad *et al.*, 2014)

Partitioning-based methods are based on divided data objects into a number of partitions, where each partition represents a cluster. K-means algorithm (MacQueen, 1967) which is most utilised clustering algorithm in literature, has a centre which is average of all points and coordinates representing the arithmetic mean.

Hierarchical-based method organised the data in a hierarchical framework depending on the medium of proximity. These proximities are obtained by the intermediate nodes. The initial cluster gradually divides into several clusters as the hierarchy continues. Hierarchical clustering methods can be bottom-up or top-down. A bottom-up clustering starts with one object for each cluster and recursively merges two or more of the most appropriate clusters. Chameleon (Karypis et al., 1999) is one of the well-known algorithm of this category. The hierarchical method's shortcoming is in their inability to undo a step.

Density-based methods separate the data object based on their regions of density, boundary and connectivity. In this method, a cluster has been defined as a connected dense component and grows in any direction that density lead to. The advantages of this method are filtering out noise (outliers) and discovering clusters of arbitrary shape. OPTICS introduced by Ankerst et al. (1999) contains information which is equivalent to the density-based clustering's corresponding to a broad range of parameter settings**.** This method for medium sized data sets can be represented graphically and for very large data sets, introduced an appropriate visualization technique. Both are suitable for interactive exploration of the intrinsic clustering structure offering additional insights into the distribution and correlation of the data.

Grid-based methods divided the space of the data objects into grids. Then, fast processing time is the main advantage of this approach. This method first employ a uniform grid to collect the regional statistic data and, then, perform the clustering on the grid, instead of the database directly. The performance of grid-based approach normally depends on the size of the grid which is usually much less than the database (Liao et al, 2004).

Finally, model-based clustering methods are based on optimisation fit between some pre-defined models and the given data. These pre-defined models are generate of mixture standard probability distributions. There are two main approaches based on the model-based methods: Natural network and statistical approaches. Natural network which self-organising maps (SOMs) (Xiao et al., 2003) is an example of these approaches, uses a set of connected input/output units, where each connection has own weight related to it. In the other side, statistical approach uses probability measures in determining the clusters.

4.2.2. *Criterion to benchmark clustering methods*

In evaluating clustering techniques for big data, there are three criteria which able to classify the strengths and weaknesses of every algorithm. Volume of data, velocity of data flows in real-time systems and variety of data types are the three major aspects in the clustering of big data which have to been considered. These three Vs (Volume, Velocity and Variety) are the core aspects and characteristics of big data which have to be taken into account when choosing an appropriate clustering algorithm.

(i) ***Volume***

Volume refers to the capability of a clustering algorithm to handle a large volume of data. The most popular classic clustering methods for their simplicity of implementation and execution of time performance, such as partitioning clustering (e.g. K-means algorithms), density based and hierarchical clustering, are limited to domain size and therefore not suitable for big data (Ayed & Halima, 2014).

(ii) **Velocity**

Velocity refers to the speed of a clustering algorithm on big data. The complexity of algorithms and run-time performance are criteria which must be considered. Time consuming algorithms which must be used several times to improve the clustering quality, are therefore not appropriate applications to run on real-time systems where new set of data samples feed into the data acquisition layer with a specified frequency.

(iii) **Variety**

Variety refers to the ability of a clustering algorithm to deal with different types of data. Most of the traditional clustering algorithms run either on numeric or categorical data. Numerical data could be either analogue or digital data and it is difficult to apply traditional clustering algorithms directly into such kinds of data.

Table 4.1 compares the five clustering methods with respects to 3Vs. Volume refers to ability of algorithm to handle high dimensional dataset and variety shows type of dataset. Furthermore, Velocity return time which algorithms needs to improve the clustering quality. Some methods must be used several times to improve quality and it takes long to handle big dataset. This comparison confirms the lack of a clustering method which is able to handle big dataset in real-time (i.e. fast) with no limitation on its data type.

Table 4.1. Comparison between five clustering method with respect to 3Vs

Method	Volume (High Dimensionality)	Variety (Type of Dataset)	Velocity (Time complexity)
Partitioned-based K-means (MacQueen, 1967)	No	Numerical	Slow
Model-based SOMs (Xiao et al., 2003),	Yes	Multivariate data	Slow
Density-based OPTICS (Ankerst et al., 1999)	No	Numerical	Slow
Hierarchical-based Chameleon (Karypis et al., 1999)	Yes	All type of data	Slow
Grid-based method (Liao et al, 2004).	No	Special data	Fast

In summary, the literature review on existing clustering algorithms for big data shows a huge amount of memory and time are required for data clustering. In addition, algorithms are too complex to implement and the majority of these algorithms handle only historical data in datasets. With early experiments with various clustering techniques reported in the literature and evaluated with the platform devised in the laboratory, showed that the Rank-Order-Clustering (ROC) method as the most suitable method of grouping in real-time. It could handle large volume and a variety of data sources with excellent efficiency and effectiveness. More importantly ROC has shown excellent capabilities to handle large data in real-time – fulfilling the most important aspects of the proposed solution.

4.3. Rank Order Clustering (ROC) technique

The Rank Order Clustering (ROC) technique introduced by King (1980a) uses matrix manipulation methods to rearrange the rows and columns of a matrix in an iterative manner. The method ultimately, and in a finite number of steps, results in a matrix form in which both the rows and columns are arranged in order of decreasing value. It is an effective algorithm to determine clusters of occurrence in block diagonal format. This approaches limited in that it is based on the assumption that groups of data will be highly similar and placed into mutually exclusive blocks. In the cluster analysis method, a group of data values are ‘similar’ according to a ‘similarity criteria’. They can either be replaced by a new value representing the group (clumping) or assigned a unique type of label (partitioning) (Groover, 2014; Ghosh& Dan, 2011; Jain, 2010).

One of the most popular usages of this method is in cellular manufacturing (CM). CM is the grouping of processes, people and machines to produce a family of products with similar manufacturing process characteristics. It is an application of a well-known philosophy called Group Technology (GT). “GT is a manufacturing philosophy in which similar parts are identified and grouped together to take advantage over their similarities and in design and production processes” (Groover, 2007).

Group Technology has been defined as the realization that it is possible to divide a large problem into manageable groups and solve it efficiently. The focus within GT is to form cells that host parts and machines with the highest relevancy. This assists designers, manufacturers and plant-layout experts to optimize production flow and movement of material. Production flow analysis and cluster analysis algorithms are the most commonly used tools to group parts and machines into cellular configurations (Groover, 2007; Vakharia, 1986).

A primary concern in CM is to determine the part families and machine cells. This is known as the cell formation (CF) problem which dissects the manufacturing systems into cells to reduce setup times, tool requirements and work-in-process inventories, and also improve product quality and productivity, shorten lead-times, and enhance the overall control of operations (Ünler&Güngör, 2009). The CF problem has long been identified as the tricky problem in grasping the concept of CM, which begins with two fundamental tasks:

(i) Machine-cell formation, where similar machines are grouped and dedicated to manufacture part-families.

(ii) Part-family construction, where parts of similar design, features, attributes and shapes are grouped and manufactured within a cell.

In the proposed event clustering technique the ROC has been used to build a cause-effect grouping of system inputs (originating from sensors/actuations) and outputs (performance indicators of the system).

4.3.1. *The ROC algorithm*

In this section a step-by-step implementation of the ROC method is explained.

Step1: Populate machine-part incident matrix (MPIM), where elements are presented as “0” or “1”. A 0 indicates no operation and a 1 indicates an active operation. Parts are arranged in columns and machines are in rows (figure 4.3).

	P1	P2	P3	P4	P5
M1	0	1	1	0	1
M2	1	0	0	1	0
M3	0	1	1	0	1

Figure 4.3. Machine-part incident matrix

Step 2: A weight for each row i and column j (in a m by n matrix) are calculated using equation 4.1(King, 1980b).

$$\text{Row i: Wi} = \sum_{K=1}^{n} a_{ik} 2^{n-k} \qquad \textbf{Equation 4.1}$$

$$\text{Column j: Wi} = \sum_{K=1}^{m} a_{kj} 2^{n-k}$$

Step 3: Read the series of 1 and 0s from left to right in the matrix as binary number 2^i(0 to i-1 number of rows) Rank the rows in order of decreasing values. In the case of a tie, rank the rows in the same order as they appear in the current matrix.

Step 4: Numbering from top to bottom, is the new order of rows the same as the rank order determined in the previous step.

Step 5: Reorder the rows in the part-machine incidence matrix by listing them in decreasing rank order (Figure 4.4).

	P1	P2	P3	P4	P5	Decimal Value	Rank
M1	0	1	1	0	1	13	2
M2	1	0	0	1	0	18	1
M3	0	1	1	0	1	13	3
	2^4	2^3	2^2	2^1	2^0		

Figure 4.4. Incident matrix row ranking

Step 6: In each column of the matrix, read the series of 1s and 0s from the top to the bottom of the binary number 2^j(0 to j-1 number of rows) rank the columns in order of decreasing value.

In the case of a tie, rank the columns in the same order as they appear in the current matrix.

Step 7: Numbering from left to right, is the current order of columns the same as the rank order determined in the previous step.

Step 8: Reorder the columns in the part-machine incidence matrix by listing them in decreasing rank order, starting with the left column (Figure 4.5).

	P1	P2	P3	P4	P5	Decimal Value
M2	1	0	0	1	0	2^2
M1	0	1	1	0	1	2^1
M3	0	1	1	0	1	2^0
Decimal Value	4	3	3	4	3	
Rank	1	3	4	1	5	

Figure 4.5. Incident matrix column raking

The final solution matrix with a block diagonal structure is depicted in figure 4.6. It shows parts assigned to machines with the largest membership index value. It proves parts 2, 3 and 5 are grouped into machines 1 and 2 and parts 1 and 4 are grouped into machine 3.

	P2	P3	P5	P1	P4
M2	1	1	1	0	0
M1	1	1	1	0	0
M3	0	0	0	1	1

Figure 4.6. Block diagonal structure matrix

4.4. EventiC method's basic parameters

EventiC defines an input and output occurrence matrix [+ -] at pre-specific time intervals. This matrix subsequently describes the relationships between causes that trigger events (trigger data) and the actual events (event data) enabling the construction of a discrete event framework for sensitivity analysis. A short description of discrete event systems, together with the definitions of trigger data and event data are provided in the following subsections. The basic parameters of the proposed Event Clustering methods are borrowed from (Tavakoli, Mousavi&Broomhead, 2013).

4.4.1. *Discrete Event Systems*

A Discrete Event System (DES) as opposed to a continuous system is defined by the disparate occurrence of events in a specified time span. This event is any change in a system state. The state of the system changes when input variable changes lead to any change in system outputs. Therefore, in DES, only the attributes that represent the occurrence of an event are considered.

4.4.2. *Trigger Data and Event Data*

Any input variable whose value transition registers an event is defined as Trigger Data (TD) in the DES. The series of data that represents the state of the system at a given time is described as Event Data (ED). Consequently, the numbers of TDs and EDs could be different, and an ED series could be impacted differently to a TD series.

$$ED :.\{ TD1, TD_2,....TD_n\} \qquad \textbf{Equation 4.2}$$

4.4.3. *Trigger Threshold (TT)*

The Trigger Threshold (TT) is a given numerical value that the values of the Trigger Data series are compared to TT like Event Threshold (ET), it is a proportion or percentage of an overall range of values of Trigger Data series over the time scale.

4.4.4. *Event Threshold (ET)*

Each transition between subsequent values of ED series is examined by Event Threshold (ET). This value is a proportion of an overall range of values of the ED series over a time scale. It is therefore expressed in the form of a percentage.

4.4.5. *Actual Value of the Data (AD)*

The series of data that represent the actual value of the data at a given time is described as actual data (AD).

4.5. Event clustering method's algorithm

In dealing with real-time event driven systems, the main logic of the proposed method is based on the assumption that changes to input variables may be interpreted as an event. Each single event or combination of events could subsequently result in a change in the system state. The proposed event clustering method describes variables and the system state as a collection of events.

The algorithm is supposed to function in real-time. The Event-Driven Incidence Matrix (EDIM) is designed based on sorting the rows for inputs and the columns for process outputs (explained in section 4.2). Incidence matrix elements can take a value of 0 and 1. The value is 1 when both or neither of the input/output event data is triggered, otherwise it is 0. This operation is similar to a logical Exclusive-NOR functionality as shown in table 4.2.

Table 4.2. Exclusive-Nor Functionality

Exclusive NOR Functionality

Input 1	Input 2	Output
0	0	1
0	1	0
1	0	0
1	1	1

4.5.1. *The assumption of the proposed method*

The Event clustering method is based on the following assumptions:

4.5.2. *Assumption 1: Delays*

The delay between EDs and the corresponding TDs is negligible and all TDs results into a specific ED (for all intent and purposes instantaneous).

4.5.3. *Assumption 2: Thresholds*

The triggers and event thresholds are a pre-specified range of signal fluctuation for every data series and determinedly the system expert which remains fix within sampling time. Thresholds are usually based on a percentage of a signal’s real value which has to meet to be assumed as an event. For example, a signal with a value of 100 units and a 1% threshold is detected as an event if its value exceeds 101 or decreases 99 units at the next analysis span (sample).

4.5.4. *Assumption 3: Homogeneity of the data series*

The event data series is assumed to be covariance stationary (The mean and variance stays constant during the analysis span).

4.5.5. *Trigger-event detection*

Equation 4.3 and 4.4 show the relationship between each event triggered by input at t and input at t-1 with respect to changes in output. Each change to the output in a given time span can be expressed as an event and the positive value of the inputs as triggers, thus output can be defined as Event Data (ED). Both $Input_t$ and $Input_{t-1}$ can be considered as Trigger Data (TD).

$$if\ (Input_t - Input_{t-1}) \geq \theta \xrightarrow{\text{Trigger}} TD_t$$ **Equation 4.3**

Equation 4.4

$$if\ (Output_t - Output_{t-1}) \geq \Psi \xrightarrow{Event} ED_t$$

Figure 4.7 shows that within each time span, (input, output) pairs are detected and used to generate the elements of the incidence matrix. The ROC method is then applied to the incidence matrices. The weighted rows and columns are clustered in the upper-left part of the EDIM. The resulting EDIM shows the ranked relevance of each input to the output. The ROC exercise leads to the cluster of the most relevant group of input event data (sensors and actuators) against output event data (plant performance indicators).

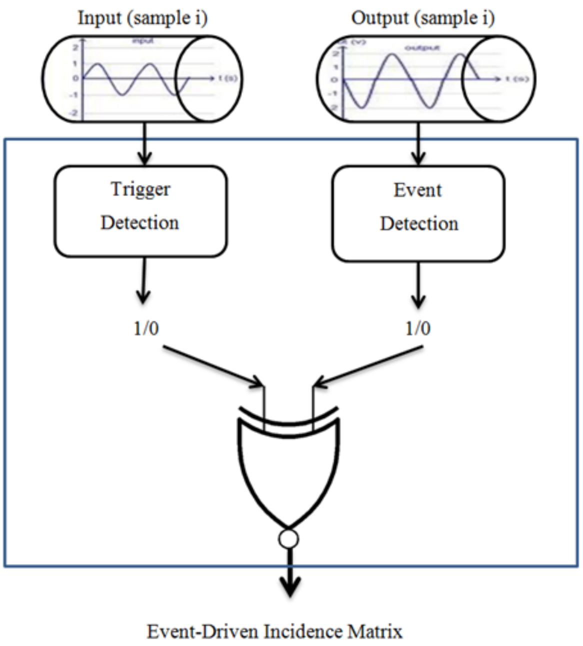


Figure 4.7. Trigger-event Detection functionality on each time scale

4.5.6. *Sample scan size*

Input and output data series scan frequency is discussed in section 4.3.1. Sample size (i.e. the number of samples which build the incident matrix) is chosen by the system expert. There is no maximum sample size for the data series but usually 250 samples are taken as minimum sample size. The data is then passed to the EventiC algorithm to build the incident matrix. Figure 4.8 shows 4 sample scans and their analysis operations in four consequent sample slots.

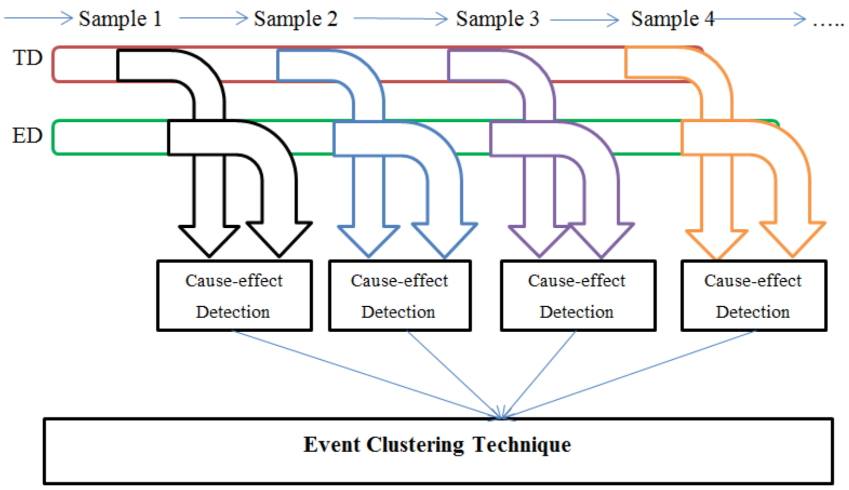


Figure 4.8. Trigger event detection functionality over whole sampling time

4.5.7. *An example to understand the implementation of the ROC algorithm*

An example here helps to advance understanding for the proposed sensitivity methods. One of the methods to detect system state transitions is to detect and track the changes that occur with input and output variables. Figure 4.9 shows the actual value of 8 inputs and 5 output variables at t_1 and t_2 sampling scan times.

In this section a step-by-step implementation of the ROC method is explained by using this example.

Step 1. Assign a 5% threshold for all inputs and outputs (an example which has to been chosen by a system expert).

Step 2. Trigger data (TD) and event data (ED) detection with respect to chosen threshold and equations 4.3 and 4.4. Figure 4.10 shows the TDs and EDs at T2. Remember that “1” is allocated to triggered and “0” to non-triggered events.

Input variables no.	t1 actual value	t2 actual value
1	10	17
2	8	13
3	12	7
4	5	5
5	70	66
6	50	50
7	10	9.5
8	11	9.2

Output variable no.	t1 actual value	t2 actual value
1	50	52
2	150	180
3	11	12
4	20	21
5	15	6

Figure 4.9. Actual value of an example Input/output at t1 and t2

Input variables no.	TDs at t2
1	1
2	1
3	1
4	0
5	1
6	0
7	0
8	1
Output variable no.	**EDs at t2**
1	0
2	1
3	1
4	0
5	1

Figure 4.10. Build TDs and EDs from actual value in table 4.2

Step 3. Build an incidence matrix from TDs and EDs in figure 4.11.

Step 4. In this section a step-by-step of the ROC method explained in section 4.2.1 is implemented on the TDs and EDs incident matrix. Figure 4.12 shows TDs 4, 6 and 7 are related to the ED1 and ED4 at the scan time.

TDs/EDs	1	2	3	4	5
1	0	1	1	0	1
2	0	1	1	0	1
3	0	1	1	0	1
4	1	0	0	1	0
5	0	1	1	0	1
6	1	0	0	1	0
7	1	0	0	1	0
8	0	1	1	0	1

Figure 4.11. TDs and EDs incident matrix

TDs/EDs	1	4	2	3	5
4	1	1	0	0	0
6	1	1	0	0	0
7	1	1	0	0	0
1	0	0	1	1	1
2	0	0	1	1	1
3	0	0	1	1	1
5	0	0	1	1	1
8	0	0	1	1	1

Figure 4.12. Final diagonal ROC matrix

4.5.8. *Average sensitivity analysis (SA) weight*

In order to find the average SA weight, all diagonal ROC matrices for whole sample numbers have to be implemented. The normalised weight of each input variable acts as the coefficient of the system outputs. Figure 4.13 shows an assumed example averaged the SA weight of TDs and EDs. For example, TD1 SA weight over ED1 is equal to 0.95 i.e. input 1 has a 95% effect over the model's output 1. An industrial case study will be reviewed in the next chapter to explain the procedure with details.

TDs/EDs	1	2	3	4	5
1	0.95	0.80	0.50	0.40	1
2	1	1	0.10	0. 15	0.40
3	0.50	1	0.32	0	0
4	0.65	0.70	1	0.62	0.85
5	0.24	0	0.90	1	1
6	0	0.75	1	0.82	0.91
7	0.95	0	0.85	0.70	0.75
8	0	0.80	1	1	0.23

Figure 4.13. Averaged SA weight of an example TDs /EDs

4.5.9. *Cut-off threshold*

Cut-Off Threshold (CT) is a mechanism to deduct the less important input variables and is in the range 0 ≤CT≤ 1(Tavakoli, Mousavi &Broomhead, 2013). For example, when CT=0.60, all inputs with an average SA weight of less than 0.6 or 60% are deducted. Figure 4.13 shows the updated format of figure 4.14 with CT=60%.

TDs/EDs	1	2	3	4	5
1	0.95	0.80			1
2	1	1			
3		1			
4	0.65	0.70	1	0.62	0.85
5			0.90	1	1
6		0.75	1	0.82	0.91
7	0.95		0.85	0.70	0.75
8		0.80	1	1	

Figure 4.14. Average SA weight of the example with CT=60%

4.5.9.1. False negative test

Besides a cut-off threshold, a false negative test is conducted to ensure that the inputs are not unnecessarily discounted. For instance, figure 4.13 proves inputs 1, 2, 3 do not have much affect over output 3 so could be filtered out. A false negative test has to be conducted by a system expert analyst to confirm these eliminations.

4.6. Event clustering sensitivity analysis design process

In order to answer research key questions which needs to overcome the shortcomings of the existing SA and clustering methods, an effective and efficient way for sensitivity analysis of data in time series is introduced in this chapter. EventiC propose a flexible data integration and clustering system architecture that helps with cost reduction of computations involved in data integration, clustering and visualization. Existing clustering algorithms for big data shows a huge amount of memory and time computation are required for data clustering. In addition, algorithms are too complex to implement and the majority of these algorithms handle only historical data in datasets.

The proposed event clustering sensitivity analysis not only using an efficient algorithm to perform sensitivity analysis, but also implementing a technique which distinguish from other automated data pattern and knowledge derivation techniques with its simplicity and speed in extracting all the available data in the system domain and converting and then processing the necessary information in near real-time. It achieves a correlation analysis throughout the analysis span and deals with too many relationships of input and output parameters. To the best knowledge of the author, no such challenge has been achieved to date. The generic nature of the EventiC solution is assumed to cover a wider range of data type which executed in real-time. Therefore, the main challenges here come from computational and time constraint.

Table 4.3 shows the proposed EventiC clustering with respect to 3Vs. It confirms that only EventiC clustering method is able to handle big dataset in real-time (i.e. fast) with no limitation on its data type, i.e. data type could be analogue, digital, binary and so on, since EventiC's algorithm build the incidence matrices regarding to variations of each data in time domain.

Table 4.3. EventiC clustering method with respect to 3Vs

Method	Volume (High Dimensionality)	Variety (Type of Dataset)	Velocity (Time complexity)
EventiC clustering method	Yes	All type of data	Fast

EventiC solution tries to overcome to these challenges in the following steps:

(i) Data collection and pre-processing
(ii) Choose an appropriate sampling rate
(iii) Define system key performance factors and indicators
(iv) Implementing ROC algorithm with a few pre-setting.
(v) Normalisation of ROC matrices
(vi) Find alternative solutions to process optimisation

All above steps will be explained in detail in chapter 5.

4.6.1. *The computational structure of EventiC sensitivity method*

Sensitivity analysis methods generally consists of two step which has been shown in figure 4.15. The first stage is to produce an iterative production of model outputs based on model's input changes (Isukapalli, 1999) and the second stage is to calculate sensitivity indices based on model's inputs and generated output values.

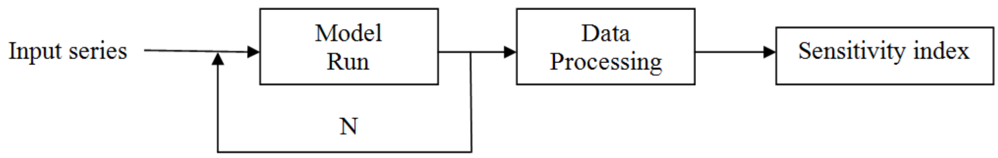


Figure 4.15. Overall structure of sensitivity analysis method

Based on access to system's model equation or using sampling-based model this structure might be modified. However, in EventiC sensitivity analysis method which is a casual statistical analysis method, method performs statistical measurements on the changes of the actual real time generated data which are produced by system input sources. Figure 4.16 shows the overall process of EventiC sensitivity analysis method.

Figure 4.16. Overall structure of EventiC method

The key feature of the proposed EventiC method is the quick filtering unimportant data that at times may overwhelm the data processing platforms. It may be safe to claim that with regard to the time domain, EventiC method may be classified as a Local Sensitivity Analysis method.

Moreover, to estimate sensitivity indices, EventiC method does not require any prior knowledge about the analytical relationship between input and output variables. EventiC, in this sense, can be considered as a Global Sensitivity Analysis method or better say model-free method.

4.7. Conclusion on Event clustering sensitivity analysis

Event clustering data grouping technique (EventiC) has been introduced in this chapter as a technique to solve sensitivity analysis problems in time-constrained complex systems. EventiC makes systems more intelligent in dealing with real-time events and provides a more accurate representation of the system, with a higher level of mathematical formalism leading to more intelligent controllers and decision making. This accurate real-time data engineer will increase precision and reduce the response time. The technology removes all the logical boundaries of isolation that exist in complex systems with the principle that every acquirable knowledge or data (input) affects the output unless proven otherwise. Therefore EventiC is not only capable of filtering unwanted data, but is capable of including information that was thought irrelevant to the behaviour of the system. This feature is unique and novel.

This technique will allow designers, engineers and system analysts who are pushed by pressing and extreme demands on performance, energy efficiency, safety, volatility of complex environments, alongside ever constraining regulations to revisit their models in use and modify them if necessary. In the next chapter more advantages of the Event Clustering technique are discussed using the framework of a cement production industry case study.

5. An Industrial Case Study

The purpose of this chapter is to introduce the proof of concepts proposed in the previous chapter by implementing the real-time unaware Event Clustering sensitivity analysis method (EventiC) and demonstrate its capability in one of the most challenging industrial industries, that of cement production.

The cement production process has a pressing requirement to become more proactive and predictive in improving the quality and efficiency of operation whilst maintaining its production rate with respect to market demands. The implementation of EventiC should lead to: (a) providing production and operation managers with the knowledge of causal relations between events that affect production. (b) Offering production engineers the necessary knowledge regarding the optimal state of production processes and machine behaviours. (c) Providing the process optimizer and decision aid system with accurate information about the relationship between key performance factors within the actual shop-floor control parameters, and finally (d) suggesting ways that can mitigate pollutant emissions and energy consumption, thus improving the kiln's energy efficiency and minimising costs. The author believes that the expected achievement can be realized through the installation of advanced EventiC monitoring and optimal control solutions specifically designed for the manufacturing process in the cement industry. Furthermore, an application of EventiC to detect unknown factors which affect the behaviour of a system will be introduced.

5.1. The cement production selection as a case study for EventiC application and other alternatives

One of the challenges in this research thesis was the selecting of an appropriate case study for the purpose of proof to the proposed EventiC technology. Cement production has been chosen for this purpose for:

(i) The economy of the cement manufacturing process is one of the most challenging compared to other industries. This challenge is due to the levels of environmental impact and regulations, energy consumption and variation of raw material during the production life cycle.

(ii) Cement's kiln is a most complex and complicated system with hundreds sensors and actuators which make it impossible to model with mathematical or analytical equations. Meanwhile, when the cement kiln is referred, it involves not only the kiln but also the adjacent systems.

(iii) Cement's kilns dataset are accessible and non-confidential in compare with other big and confidential industries.

(iv) Due to cement important role in infrastructure development, a number of research projects are conducted to establishing a more efficient and effective production method in cement plants.

These reasons turn cement's kiln a suitable experiment to our EventiC, a complex system with a high number of system's input (sensor and actuator) and a few number of key performance factors (system's output). However, the application of EventiC is not limited to cement industry and could be applied to any industries with big and complex environments, for examples vehicular transport (aircraft, automobile and other vessels), financial market or Internet of Things. It also could be applied in bio-informatics, for instance, in gene expression datasets which consists of big heterogeneous gene dataset, sophisticated computational and clustering/visualization methods (Basel et al., 2013).

The cement kiln control begins with using a fuzzy controllers as one of the first successful applications of the fuzzy controller in industry. Holmblad and Ostergaard used the first fuzzy controller for a cement kiln control in 1978. They saw that the results were much better than when the kiln was directly controlled by humans (Wang, 1994). Nowadays, the cases of using the fuzzy logic controllers for controlling the cement kilns have been increased. This is based on this fact that the fuzzy logic controllers do not need an accurate model of the plant. By fuzzy logic controller, a remarkable improvement of the cement quality and a decline in the production expenses has been achieved. Several first designs of such controllers have been proposed and/or implemented in (Devedzic, 1995), (Bo *et al*, 1997), (Ruby, 1997) and (Tayel *et al*, 1997), which have been designed based on the knowledge of the operators. In next chapter, the auto extraction of a fuzzy control system for industrial processes using of EventiC application will be introduced. This integration could be replaced to the proposed solution in (Mendes et al., 2011) which uses GA to manipulate the parameter selection of the fuzzy system.

5.2. The cement production industry

The main objective of this section is to gain a substantial understanding of the cement kiln process in order to optimize and control the functionalities of kilns. The EventiC algorithm will be integrated with the kiln process in the next section.

The cement production industry plays an important role in a country's infrastructure, thus a number of research projects have been carried out towards establishing a more efficient and effective production method in cement plants. Furthermore, due to the consumption of high energy within the production process, there are intensive efforts to mitigate emissions and costs. Analysis of the current situation in the cement industry leads to the outcome that sustainability, energy efficiency, cost reduction and CO_2 emissions mitigation are the basic principles of the industry.

A major stage of cement production is the formation of clinker, a process executed in the kiln (Conesa *et al.*, 2008). Therefore, an area that will be carefully examined is the optimization of energy consumption in the kiln process.

Energy consumption, its impact on the environment and the cost of production are significant in the cement industry. The goal is therefore to examine different scenarios to both reduce the effects of emissions produced by the cement industry and increase kiln productivity. As much as productivity of a cement plant is a market objective, the pressure for emissions reduction is more intensive, because CO_2 emitted by cement production is a major contributor to air pollution.

On a global scale, the construction industry is one of the most important and highly competitive markets. As a key ingredient of concrete, cement holds together roads, bridges, buildings and other structures (Battelle, 2002). The cement production process has several stages, from the extraction of raw materials to the storage and delivery of the final product. The current analysis will focus mainly in the procedure that takes place in the kiln.

The kiln process is highly intensive in terms of energy consumption and thermal demand as the raw material reaches relatively high temperature levels, of about 1350-1550°C. Due to its high volume of production, the cement industry is one of the major contributors of Green House Gas (GHG) emissions. The amount of pollutants is expected to increase at a high rate, as the demand for cement production is projected to increase considerably within the next few years. However, the existing methods for reducing CO_2 emissions do not seem capable of offsetting such growth (Damineli et al., 2010). In order to create an efficient and environmentally sustainable industry, serious attention needs to be given to the optimization of the cement production processes.

5.2.1. *Cement kiln process*

Figure 5.1 represents a general overview of the cement production process (IEA & WBCSD, 2009).

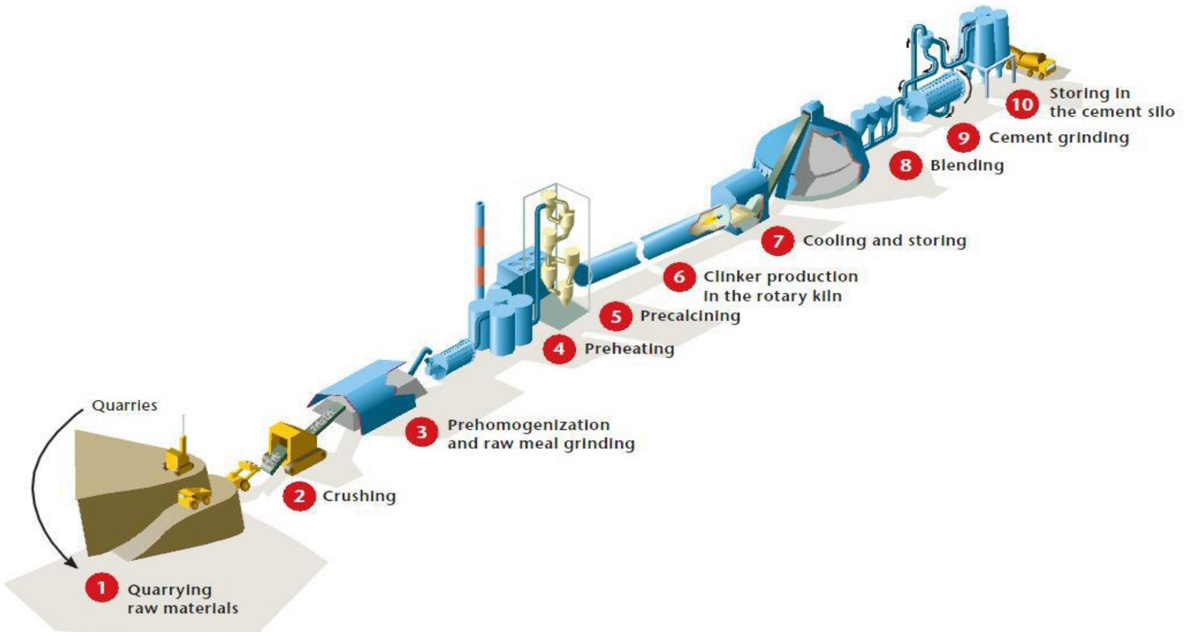


Figure 5.1. Cement production overview

The cement manufacturing process is divided into five main stages that are outlined below (Battelle, 2002; Cement sustainability initiative, 2015):

1. Extraction of raw material

Calcium carbonate, silica, alumina and iron ore are raw materials needed for the formation of cement. These materials are gathered from limestone rock, chalk, shale or clay quarries, either by extraction or through blasting.

2. Raw mill

After the completion of the mining process the raw materials are prepared prior to thermal processing. In order to achieve a better efficiency within the kiln process, the materials need to be milled to the relevant size, mixed in the correct proportions and then dried. Once dried, this homogenised raw material is stored in large silos as a powder called flour. Older plants utilise a wet process to prepare the raw material whereby the materials are mixed with water to form slurry.

3. Kiln process

The raw material is moved into the kiln through a process called 'sintering'. The burner produces heat to maintain a material temperature of between1350-1550 °C. The resultant material is known as clinker, which is a mixture that contains hydraulic calcium silicates.

4. Cement mill

The clinker is stored until required for the final step of creating the end product, and if needed, it is ground to a fine powder and mixed with gypsum and other additives. These 'additions' can affect the cement's properties, altering its permeability, workability and resistance to sulphates. Depending on the required final product the amount of clinker may range between 30% and 95%.

5. Shipping

Cement, as a final product, can be stored in silos and then be transported in bulk or in bags as required.

5.2.2. *Kiln system*

The kiln system involves not only the kiln, but also the pre-heater and the cooler. A simplified diagram of the kiln system and its subsystems is shown in Figure 5.2:

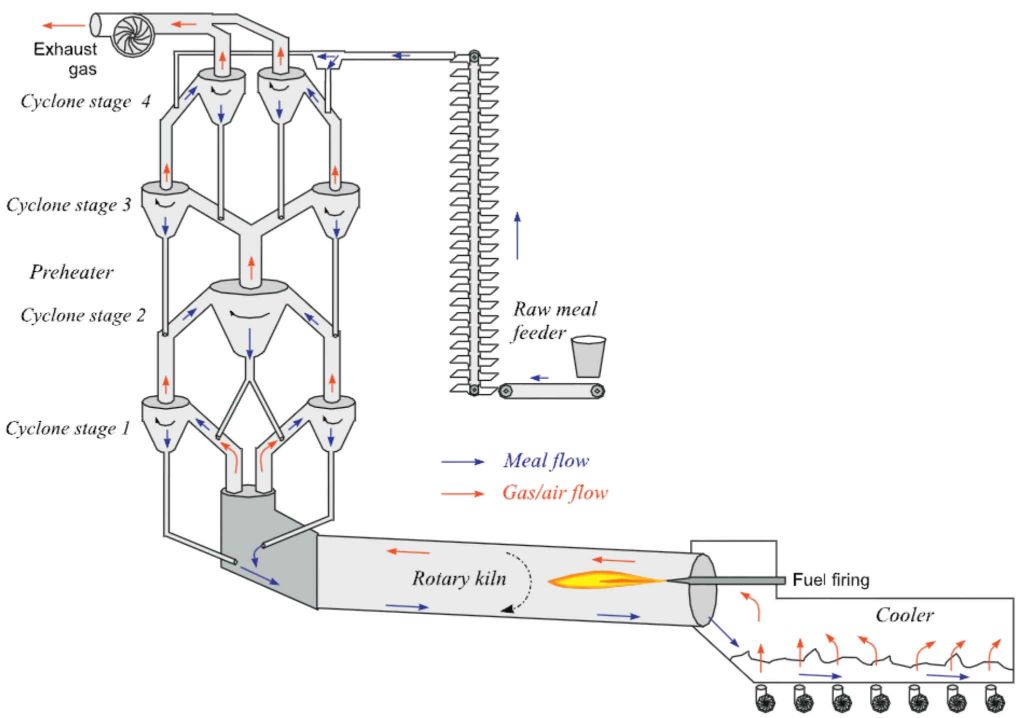


Figure 5.2. The kiln system diagram

The kiln system consists of the following three parts (Mousavi et al., 2011):

- **Pre-heater tower**

In order to be heated before entering the kiln, the raw material passes through the pre-heater which is a tower with a series of cyclones (usually there are four cyclones stages, as shown in figure 5.2). The high temperature that is required for this process (around 900°C) is reached through a combination of hot air coming from the kiln and the direct injection of burning material with the aid of a calciner (Fidaros*et al.*, 2007).

- **Kiln**

The main process of cement production takes place in the kiln, where clinker is produced. The kiln process is highly energy intensive and accounts for about 90% of the total energy consumption (Kabir *et al.*, 2010). The two main processes executed in the kiln furnace are calcination and sintering. These two procedures are very energy demanding, as they both require high levels of heat transfer.

Calcination is a chemical conversion process in which calcium carbonate ($CaCO_3$, limestone) is converted to lime (CaO), the primary component of cement. $CaCO_3$ is decomposed at about 800- 900°C producing CaO and emitting CO_2 as a by-product. This process is described by the following general reaction in equation 5.1.

$$CaCO_3 + heat \dashrightarrow CaO + CO_2$$ **Equation 5.1**

After the calcination, CaO reacts with silica, alumina and ferrous oxide in order to form the clinker, at a temperature of 1350-1550°C.

Sintering is the process where the raw material is heated in order to form the clinker. The material requires a temperature of about 1450°C, which is reached by fuel combustion. The fossil fuels that are used emit large amounts of CO_2, thus researchers have tried to find alternatives to replace the traditionally used fuels (i.e. coal, petroleum coke, fuel oil, natural gas).

- **Clinker cooler**

When clinker exits the kiln it has a temperature that makes it impossible to handle. Therefore the main objective of the cooling system is to cool the clinker so that it is able to be handled and stored. However, the process needs to be rapid, for slow cooling makes the material unstable in lower temperatures (below 1250°C), meaning it could revert to belite. This frees CaO, which is undesirable with regards to clinker quality. This means that the cooling process needs to be done precisely. Finally, the excess heat is reused in the form of hot air in the preheating stage.

5.2.3. *The kiln's key performance factors and indicators*

One of the most important tasks in order to be able to examine the operation of each system and provide applicable solutions for the improvement of the production process is to identify the key performance factors and indicators of the production process.

5.2.3.1. Definition of key performance factors and indicators

Key performance factors (KPF) are factors of a system which are normally determined by experts in that system and are related to their influence on the profit margins of the system. For instance, KPF in a manufacturing system could be categorized as: levels of customization, productivity, resource utilization, efficiency and inventory management. These factors can be defined loosely and in general terms (i.e. subjective terms) which then need to be broken down into key performance indicators (KPI). These indicators are objective and normally metricized. For example, Mousavi *et al*. (2006) classified the performance of production system costs using five key factors:

1. Resource Utilisation (RU)
2. Customer Satisfaction (CS)
3. Productivity (PR)
4. Inventory Management (IM)
5. Efficiency (EF)

These factors can then be prioritized according to the system’s preferences and requirements. Their importance may vary from one market to another, or between different competitive strategies.

Figure 5.3 demonstrates a typical example of the relationship between a key performance factor (in this example, customization) and key performance indicators (Mousavi et al., 2006).

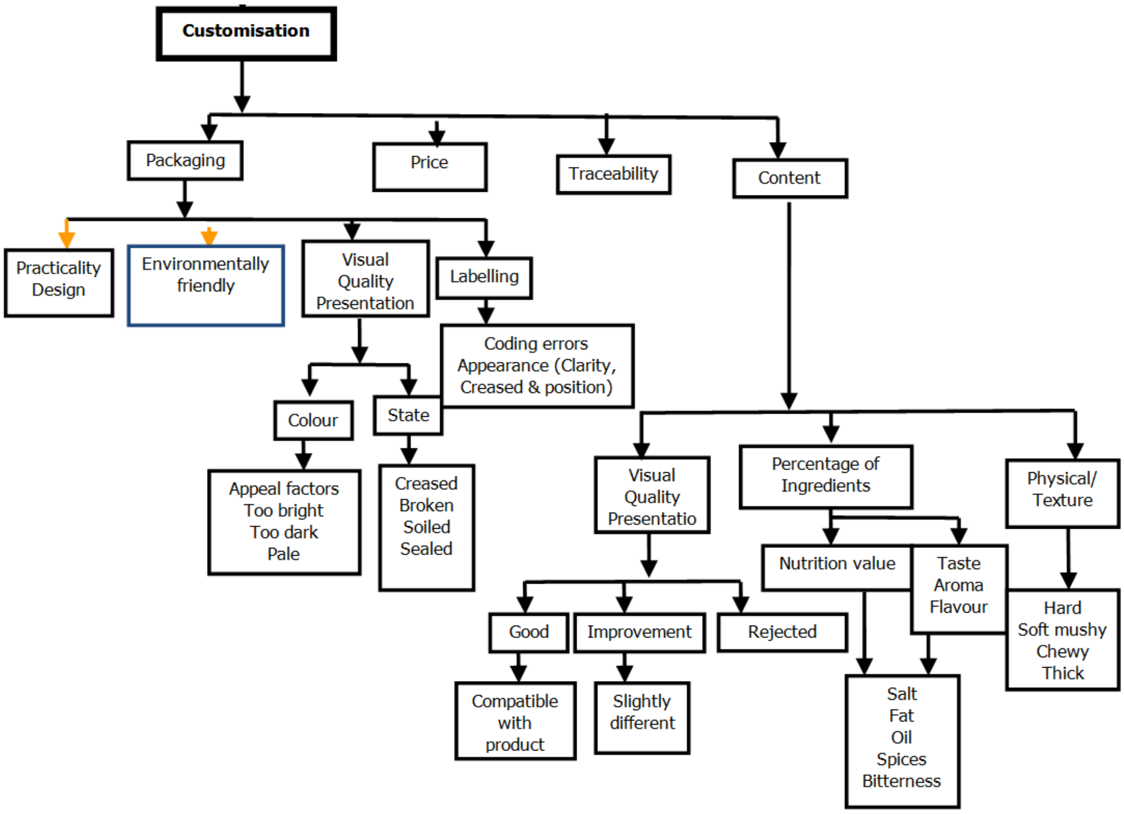


Figure 5.3. KPIs of the customization factor in a manufacturing system (Mousavi *et al.*, 2006)

5.2.3.2. KPF and KPI in cement production

The nature of the cement industry is complex, dynamic and competitive. However, managers of systems have to measure and monitor the most important performance factors and indicators that will improve the cement production. Chan (2004) suggested that when deciding about the suitable KPFs and KPIs the following factors should be taken into account:

- KPFs and KPIs are general indicators of performance that relate to significant aspects of outputs.
- Only a limited and manageable number of KPFs and KPIs are possible to handle.
- Data collection must be made as simple as possible.

The key factors that influence kiln systems are multiple and in some cases interrelated. The most important ones are divided into two categories: the on-line and off-line parameters.

Values for the online parameters can be gathered by real-time data acquisition. For instance, the variables in the cooler are the flow at the fan inlet, motor amperage, fan discharge pressure and the temperature of recovered air. The variables in the kiln are the pressure at the kiln hood, the amount of fuel, air pressure and flow in the burner, kiln speed and motor amperage, plus gas temperature and pressure in the chamber. In the cyclone tower, inlet and exit pressure and temperature of the gas, material temperature and gas analyser are the major variables. Finally, in the exhaust fan the most important parameters are inlet speed, pressure and temperature, and amperage of the exhaust fan.

In contrast, off-line parameters are the results of laboratory data analysis about clinker, hot meal in the kiln inlet chamber, raw material from the first cyclone and raw meal mill discharge.

In cement production, the process that takes place in the kiln is not only important, but also somewhat complicated due to the large number of interconnected variables. Therefore, a major objective of this chapter is to focus on the kiln system process and implement EventiC on a kiln's SCADA and then take into account different scenarios that will optimize a kiln's functionalities.

The cement production key performance factors are classified as following:

A. Environmental impact and energy consumption

Cement manufacturing processes have several direct impacts on the environment. The major factors that are connected with cement production and affect the environment are the following (Cembureau, 1999):

1. Dust (stack emissions and fugitive sources)
2. Gaseous atmospheric emissions (NO_X, SO_2, CO_2, VOC and others)
3. Other emissions (noise and vibrations, odour, process water, production waste)
4. Resources consumption (energy, raw materials)

Fuel oil, natural gas, petroleum coke, coal, wood pellets and tyres are all used as energy sources in the cement furnace. Raw materials and fuels have large concentrations of sulphur. The high temperatures that are needed for raw materials' processing cause them to release considerable amounts of SO_2. In addition, the high temperature in the combustion process and the chemical composition of materials result in the formation of NO_x. Moreover, during calcination process, the organic carbon of raw materials causes the emission of CO, CO_2 and VOC. Finally, in the past, the combustion of fossil fuels produced high volumes of CO_2emissions. However, successful efforts have been more recently been made to reduce the amount of CO_2 emitted by fossil fuels.

On a global scale, it is estimated that the cement industry is responsible for about 5% of man-made CO_2 emissions (Naranjo et al., 2011). During the cement production process, the following three factors affect CO_2 emissions:

- Lime production
- Cement kiln dust
- Fuel combustion

In the kiln process, the measurement of energy consumption is a representative indicator of the system’s performance. The mixture needs to be heated at very high temperatures, thus the kiln requires intense energy consumption.

Literature reviews in the cement industry show a linear relationship between kiln temperature and energy consumption (Dimitrios& Evangelopoulos, 2012). Figure 5.4 shows their relationship and methodology for the calculation of energy consumption. It should be mentioned that the ideal kiln temperature is between 1350°C to 1550°C.

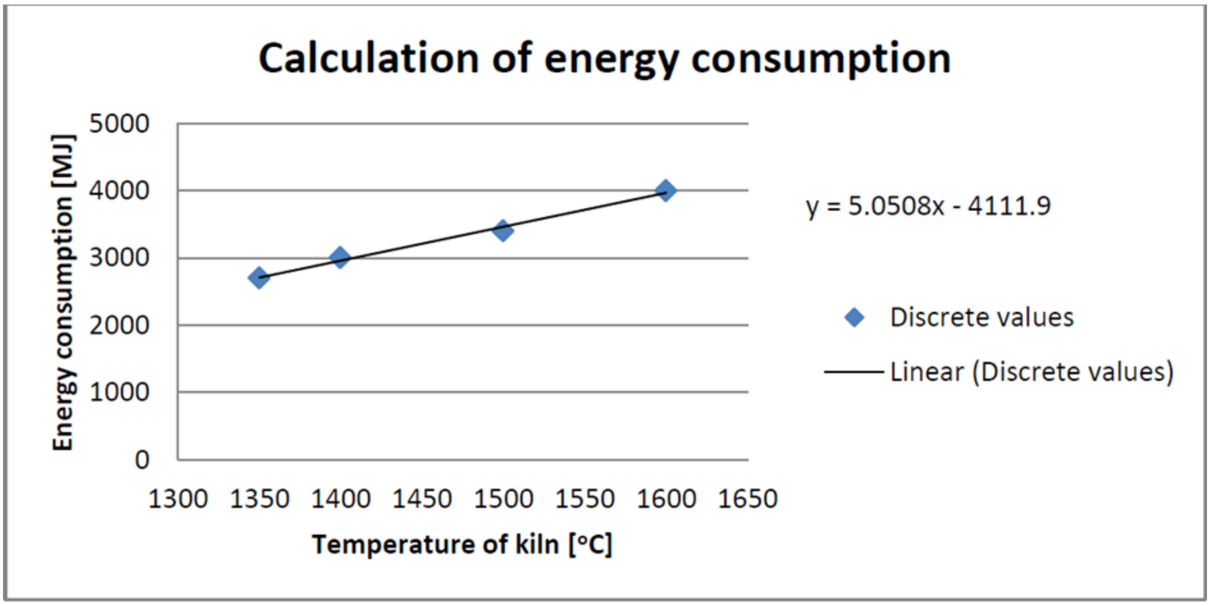


Figure 5.4. Calculation of energy consumption
Source: (Dimitrios, 2012)

Moreover, assuming that 90% of the total energy consumption is consumed in the kiln, the model calculates the total energy consumption for a given amount of cement produced. The calculation formula is as follows:

Total energy consumption = **Equation 5.2**

1.1 * kiln energy consumption * tonnes of cement produced

B. Efficiency of cement kiln production

One of the most challenging elements of the cement manufacturing process is energy efficiency. The heating process performed in the kiln is highly energy intensive with an average fuel consumption of 2.95 GJ per ton of cement produced for well-equipped advanced kilns (Engin& Ari, 2005). The energy efficiency of kilns is affected by several factors. There is a wide range of different existing technologies that give numerous options to the manufacturing industries. However, the longevity of kilns, which is around 50 years, makes it difficult for a company to change its existing technologies with new ones very frequently. In addition, it is hard to evaluate the cost effectiveness of a kiln replacement as it involves serious consideration of many different factors, such as installation costs, maintenance, suitability of the characteristics for meeting customers' demands and compatibility with the existing designed processes (Saidur et al., 2011).

C. Cement Quality

In the cement industry, the term 'quality' with reference to cement does not necessarily refer to 'high quality' products as there are different types of cement for different uses. Thus, quality does not only derive from a system's performance, but is also affected by demands and reflected in customers' satisfaction. In order to represent the different needs for outputs, three types of cement are assigned: high quality, medium quality and low quality cement. The parameter that controls this categorisation is the existence of free lime in the mixture.

Table 5.1 indicates the outputs of cement quality with reference to three different kiln temperatures.

Table 5.1. Type of cement quality VS kiln temperature (Yao*et al*., 2010)

Type of cement quality	Average Kiln temperature
high quality cement	1550 °C
medium quality cement	1450 °C
low quality cement	1350 °C

As shown in Table 5.1, kiln temperature is directly related to the quality of the final product. Measuring and controlling the temperature of a kiln is a complex task which is affected by several different factors. According to a relevant study, the model of a rotary kiln temperature control system has a high order time, and nonlinear, coupled, multivariable, parametric time-varying characteristics, so it is difficult to obtain an effective analytical model (Yao *et al.*, 2010).

D. Productivity (production rate)

Productivity can be associated with yield. Yield is defined as the variation in percentage of the final product in output over input feed. The significance of the kiln function and its contribution to the whole process of cement production sets the kiln temperature as one of the major performance indicators of the kiln. Along with kiln temperature, raw material selection is an essential input that conducts the productivity of the kiln (Saidur *et al.*, 2011). There are other indicators which effect productivity such as kiln feed rate, kiln fans and the kiln's inner motor torque. All of these will be discussed in the next section.

5.3. A case study for EventiC application

This section tries to map the design requirements that you have now explained better in Chapter 4 to an obvious EventiC's design process.

With a case study, raw information sourced from the existing SCADA of a cement plant has been used. The data acquisition of the SCADA system consists of 196 sensors and actuators that are used to monitor the kiln and its surrounding equipment. Appendix A includes the kiln schematic, names of sensors/names of actuators, and all the raw data taken from the SCADA of our industrial partner. EventiC's design process are drawn in table 5.2.

Table 5.2. EventiC's design requirement and design process

Design requirements	EventiC's design process
Systems with no limited boundary?	Data collection stage and quick response ROC algorithm
Volatile system?	Data sampling frequency stage
Find appropriate corrective actions to optimal functionality	kiln operation's key performance factors and indicators definition
How to detect real-time events?	Build event-driven incident matrices
Big systems which need quick responses and high computational?	Quick response ROC algorithm
How to improve data quality?	Filter out irrelevant data at the dataset

These steps will be explained in details in the following of this chapter.

5.3.1. *5.2.1. Data collection and pre-processing*

As explained above, the raw data source is the existing cement plant SCADA system. In terms of data acquisition the SCADA system consists of 196 sensors and actuators that provide EventiC with input event data (TD) that monitors and controls the kiln and its peripheral equipment. Data collection was conducted over a one month production period, collecting approximately 43,000 data samples. After collecting data from the rotary kiln, the data was pre-treated for identification processes. One of the main reasons for data pre-treatment is high frequency noise and spikes on the main raw data, for sometimes immeasurable disturbances occur and take the system out of its normal operating points. To solve the problem of high frequency noise and spikes, pre-processing methods (applying proper filters) is conducted (Nelles, 2001) to obtain better data for identification process.

5.3.2. *Sampling frequency*

During the sampling of continuous signals, some information might be annihilated. Therefore, it is essential to choose a proper sampling frequency which does not interfere with the control system. Zhu (2001, p.57) proposed three methods for the sampling of continuous signals in system identification.

Smallest time constant	$T_s=\tau\ min\ /3$	**Equation 5.3**
Bandwidth	$f_s=10f_0$	**Equation 5.4**
Settling Time	$T_s=T_{st}/20$ to $T_{st}/100$	**Equation 5.5**

In our experimentation, in order to obtain sampling frequency, the study of output signals shows the smallest time constant of the system is three minutes. Then, according to equation 5.3, the data sampling rate of the system is set at 1 per minute. The event modelling process was conducted over one month of production (i.e. approximately 43,000 data samples).

5.3.3. *Implementation of EventiC on kiln operation*

The EventiC algorithm resides over the data tables of the factory control and monitoring system, where the cause-effect relationship between triggers of inputs (TDs) and triggers of events (EDs) are measured. The EventiC algorithm generates a new output matrix at each system scan rate (a scan per minute).

The output event data (EDs) is collected from the sensors and counters that measure the production rate (the output of the kiln is defined as the volume of the satisfactory product leaving the kiln), energy consumption (the kiln temperature is directly related to energy consumption) and CO_2 emission.

In the following sections each of these key performances will be evaluated.

5.3.3.1 Production rate

Product leaving the kiln is one of the existing 196 sensors and actuators in the kiln. As explained, the data sampling rate of the system is set at 1 per minute. The event clustering process was conducted over a month production period (producing approximately 43,000 data samples). The EventiC algorithm resides over the data tables of the factory control and monitoring system, where the cause-effect relationship between triggers of data (changes in sensors outputs) and events (changes in system outputs) are measured. A new output matrix is generated on each scan of the system; the normalised weight of each input variable acts as the coefficient of the system outputs (Average SA weight). Figure 5.5 shows the EventiC application of all 196 inputs over the kiln's production rate as a model's performance output.

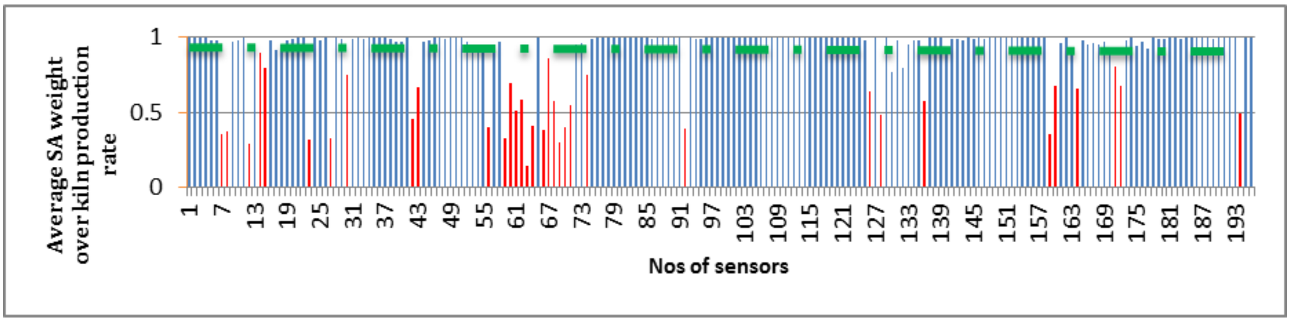


Figure 5.5. The kiln's production rate SA with respect to 196 inputs over 1 month sampling snapshots with 90% cut-off threshold

5.3.3.1. The Cut-off (CT) threshold

As mentioned in section 4.5.6, the CT is a mechanism to deduct the less important input variables and is in the range 0 ≤CT≤ 1. A false negative test is conducted to ensure that the inputs are not unnecessarily discounted. Figure 5.6 shows the percentage of filtered TDs over the plant production rate with regard to different CT and the ratio of false negative in the experiment.

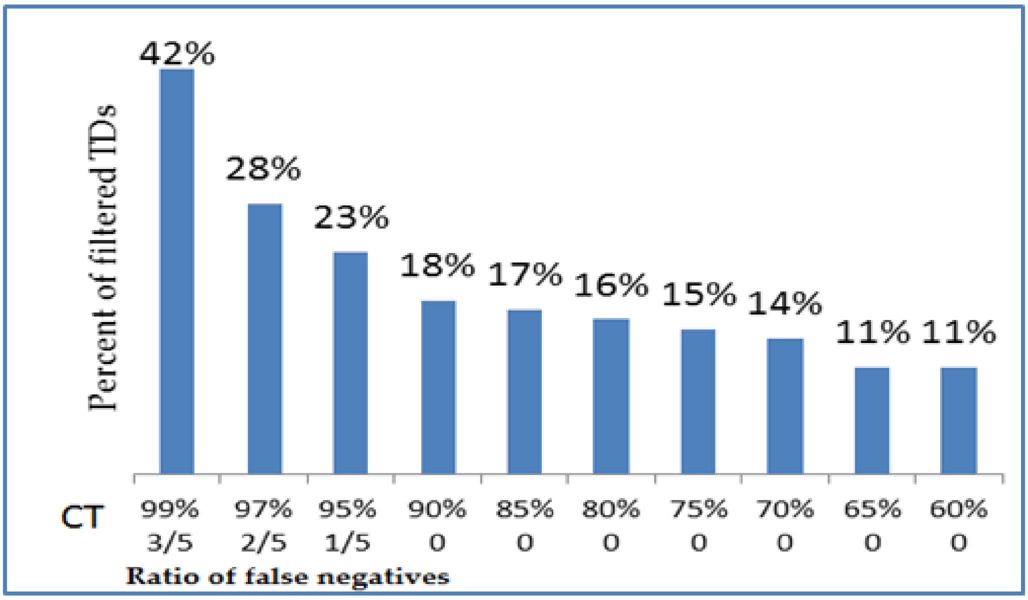


Figure 5.6. Percentage of filtered TDs per CT and the ratio of false negatives over the production rate.

In the experiment, an arbitrary CT of 90% (horizontal dashed green line in figure 5.5) has been chosen to eliminate the false negatives. Based on this CT, 18% of triggered inputs (36 TDs) have been filtered out (The red bars in figure 5.5).Thus the remaining 82% of input sensors represent all the necessary input variables that affect the state of the kiln.

In table 5.3, nine of the event data inputs and their corresponding weights on the kiln's production rate are listed. The input variables in bold represent the inputs that have the highest impact on the kiln's production rate. The weighting mechanism mentioned earlier is based on the number of times the input-output event data coincided during the analysis span comprising 43,000 data points. Regarding the inputs' SA weight, kiln temperature, I/h return, kiln fan and motors have influence over the kiln's production rate and the other of the kiln's sensors and actuators have effects over other key performance indicators. Appendix B includes all 196 sensors and their sensitivity weight over the kiln's production rate.

Therefore, following the event modelling concept explained in chapter four, the key performance input vector for the kiln's production rate which is $V = [x_1, \dots x_9]$is deducted to the more reliable input vector $V = [x_1, \dots x_5]$. This input vector dimension reduction prepares much more reliable and effective state models of the kiln's system.

Table 5.3. Averaged SA weight of a selected number of the kiln's input data over the kiln's production rate

Input Name	Sensitivity Level of kiln Production Rate	Subjective Importance level with CT=90%
Kiln temperature	**92 %**	**High**
CO output	63 %	Low
I/h return in kiln	**90%**	**High**
Kiln fan	**98%**	**High**
CO_2 output	**97%**	**High**
Motors pulls material from kiln	**92%**	**High**
Injected O2 to kiln	37%	Low
Injected NO2 to kiln	54%	Low
Injected SO2 to kiln	36%	Low

5.3.3.2. Kiln's CO_2 emission

In this section, the kiln's CO_2 emission as a major cement industry's environmental pollutant is considered as a model's key performance (output) and then the EventiC algorithm was conducted over it for 1 month sampling snapshots. Figure 5.7 illustrates the normalised weight of each input variables over the kiln's CO_2 emission.

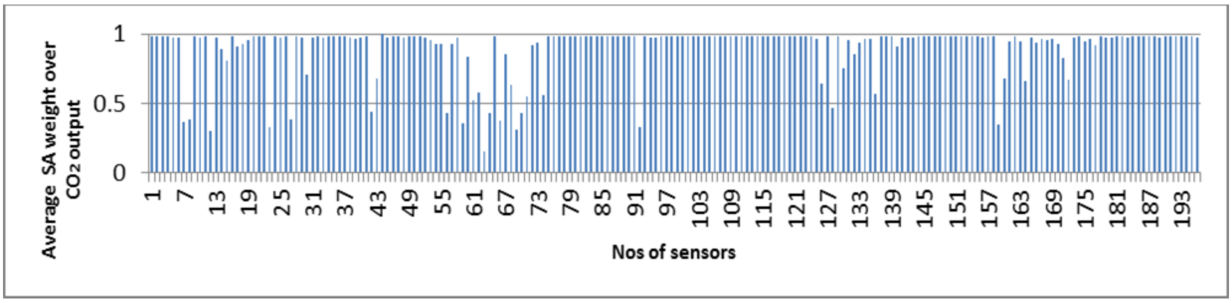


Figure 5.7. The kiln's CO_2 emission SA with respect to 196 inputs over 1 month sampling snapshots.

In table 5.4, 9 of the event data inputs (9 out of 196 existing sensors and actuators) and their corresponding weights over CO_2 emissions are listed. It confirms that 4 of these inputs (with CT=90%) have a considerable contribution on environmental pollutants and others have less effect with different effect weights. Thus, a key performance input vector for the kiln's CO_2emissions deducted to $V = [x_1, \ldots x_4]$ provide a much more reliable and effective system state model for the kiln.

Table 5.4. Averaged SA weight of some selective kiln's input data over CO_2emission

Input Name	Sensitivity Level of Kiln CO_2 output
Kiln output	**97%**
Kiln temperature	**92 %**
CO output	64 %
I/h return in kiln	56%
Kiln fan	**97%**
Motors pulls material from kiln	**97%**
Injected O_2 to kiln	37%
Injected NO_x to kiln	54%
Injected SO_2 to kiln	36%

5.3.3.3. Kiln's energy consumption

As shown in figure 5.4 and with equation 5.2, the kiln's energy consumption is directly related to kiln temperature. Thus in this section kiln temperature is assumed as a model's output instead of kiln's energy consumption. Figure 5.8 shows the EventiC algorithm output.

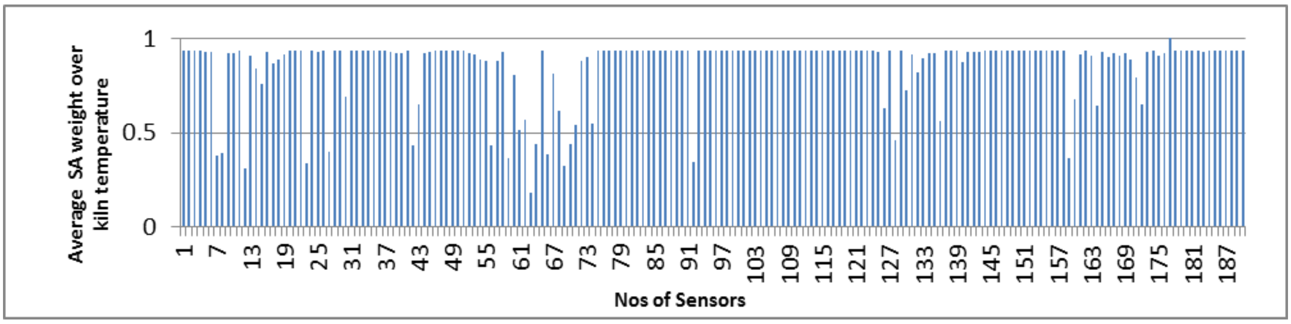


Figure 5.8. The kiln's temperature SA with respect to 196 inputs over 1 month sampling snapshots.

In table 5.5 nine out of the 196 existing sensors and their corresponding weights over kiln temperature are listed. It shows that 5 of these inputs make a considerable contribution on the kiln's energy consumption and others have less effect with different effect weightings.

Table 5.5. Averaged SA weight of some selective kiln input data over kiln temperature

Input Name	Sensitivity Level of kiln temperature
Kiln output	**93 %**
CO output	**90 %**
I/h return in Kiln	55%
Kiln fan	**93%**
CO_2 output	**91%**
Motors pulls material from kiln	**93%**
Injected O_2 to Kiln	38%
Injected NOx to Kiln	53%
Injected SO_2 to kiln	37%

5.4. EventiC and real-time plant control

In this section we will discuss how EventiC becomes a stability/optimisation tool in process control, or can alternatively provide high quality data for higher level optimization/autonomous/intelligent systems to conduct their tasks.

One of the major features of the proposed EventiC is the grouping and clustering series of relevant inputs to system outputs. In essence, the clustering mechanism in the EventiC bundles inputs and outputs together using an incident matrix in real-time (see figure 4.5). This capability allows the grouping of key system input parameters to system output parameters.

For the purpose of optimization, or returning a system back to stability from an excitation that caused instability, to quickly identify the source of excitation (i.e. input) and its level of impact, would allow the controller (automatic or human interference) to return the state of the input back into the normal state (corrective measure), or if the system is running an optimization process based on some key performance indicator. As an example, imagine the energy consumption of the kiln running at a steady state ranging between 2-2.5 KW/h. The optimal energy consumption would be the minimum value, whilst the kiln retains the same performance specifications. EventiC will be able to advise on the most economical way to maintain the settings of actuators to keep the kiln energy consumption at 2.1 KW/h at all times. For example, let the state of a system, demonstrate a period of instability caused by excitation ($0 \leq t \leq T_s$) and then reach a steady-state stability as shown in figure 5.9.

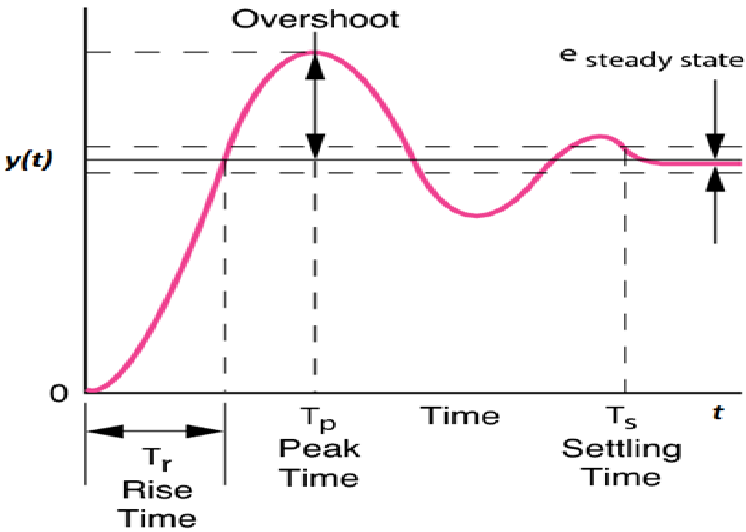


Figure 5.9. System state in time

A system in a steady-state (e.g. kiln production) needs to remain in that condition. However, there is the possibility that an internal cause (within the kiln) or an external cause (outside the kiln) provokes an excitation that destabilises the kiln. At this juncture the most important information is to identify the source of excitation (real-time) and quickly identify the impact on the system. Knowing the clusters of input variables and their impact on returning the system back to a stable condition is vital.

This section tries to suggest ways that can improve the functionality of kilns, reduce their energy consumption, mitigate the environmental impacts of their processes, and at the same time, keep the kiln efficiency at the required levels. The deployment of different scenarios will show how the parameters that are examined can be beneficial for the industry, whilst maintaining the required standards of the cement produced.

As discussed in the previous section, the proposed event clustering technique is used as an input variable selection technique to provide high-quality input data to the higher-level optimization, autonomous or intelligent systems in a short period of time. Using the example of the cement plant and kiln operations, it can be demonstrated how this technique becomes a stability/optimization tool for process control.

One fascinating observation whilst conducting the experiments was the realization that reaching a specific output (i.e. production rate) may not necessarily be the result of a single cluster of system inputs (i.e. system settings). Figure 5.10 shows the five highest plant production rates at 9.68-9.76 ton/hour over the analysis span (i.e. one month production) in our experiment.

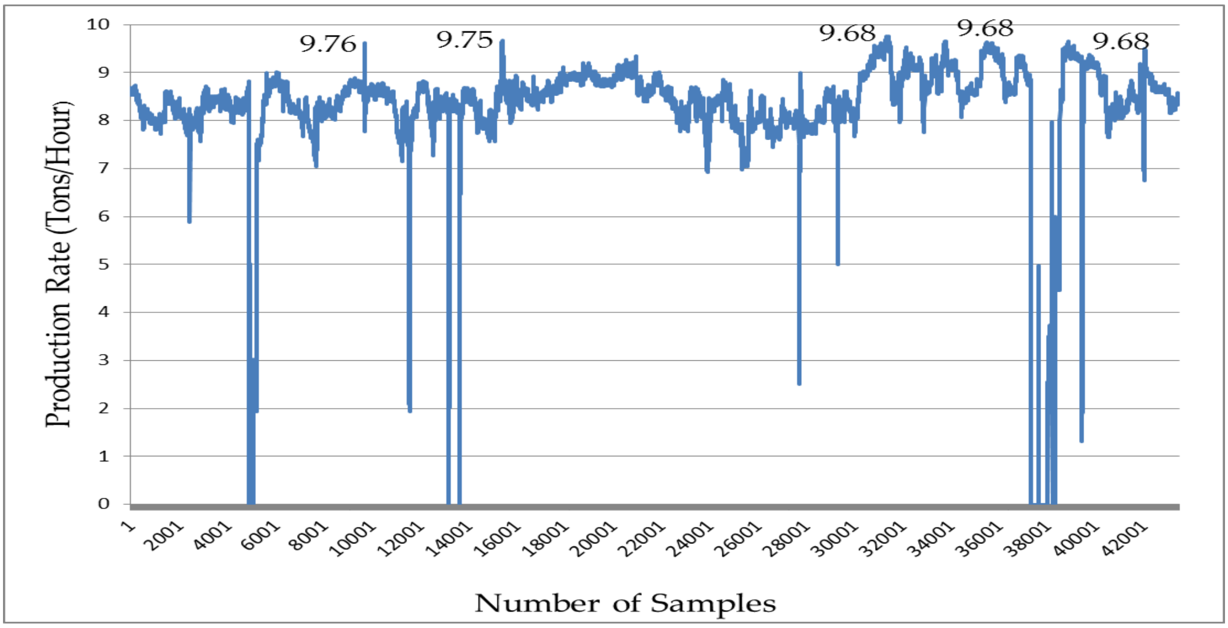


Figure 5.10. Actual value of production rate and its maximum points over 1 month sampling snapshots.

These five maximum plant production rates and their corresponding clusters of inputs (i.e. plant settings) are listed in table 5.6. The information shows that it is possible to maintain the highest production rate using five practical alternative settings. This was not possible before. By referring to the data series, and with respect to simple production economics, it shows that the control system can be set at the lowest possible cost (i.e. the least energy consumptive rate or at the minimum CO_2 emissions rate) at maximum production rate. In this case solution 1 (row 1) demonstrates a solution that keeps the production rate at maximum and retains the lowest temperature in the kiln (less energy consumption). Solution 2 (row 2) represents the minimum CO_2 emission, whilst maintaining the maximum production rate.

Table 5.6. Five alternative clusters of input variables and maximum kiln production rates

	Kiln Production Rate (Tons /Hour)	Kiln Temperature (degree Celsius)	I/h return in Kiln (%)	Kiln fan (on/off)	CO2 output (%)	Motors pulls material from Kiln(on/off)
Solution 1: Minimum Kiln Temperature	9.76	1274	18.34	1	15.47	1
	9.75	1355	13.42	1	15.50	1
Solution 2: Minimum CO2 emission	9.68	1368.37	17.42	1	14.52	1
	9.68	1369.65	18.15	1	15.55	1
	9.68	1371.7	15.05	1	15.47	1

Tables 5.7 and 5.8 provide the information with regards to the input variable clusters that determine the energy consumption and CO_2 emission of the kiln.

Table 5.7. Five alternative clusters of input variables in minimum kiln's energy consumption

	Kiln Production Rate (Tons /Hour)	Kiln Temperature (degree Celsius)	CO output (mg /m^3)	Kiln fan (on/off)	CO2 output (%)	Motors pulls material from Kiln(on/off)
	2.8	1053.3	98.39	1	2.69	1
	3.47	1053.3	80.08	1	3.34	1
Solution 1: Minimum CO2 emission	5.00	1053.3	81.79	1	0.111	1
	8.45	1053.3	799.84	1	12.19	1
Solution 2: Maximum kiln productivity	8.97	1053.3	159.6	1	12.07	1

Table 5.8. Five alternative clusters of input variables in minimum kiln's CO_2 emission

	Kiln Production Rate (Tons /Hour)	Kiln Temperature (degree Celsius)	CO output (mg /m^3)	Kiln fan (on/off)	CO2 output (%)	Motors pulls material from Kiln(on/off)
Solution 1: Minimum Kiln temperature	8.10	1283.3	383.28	1	0.037	1
	8.57	1371.37	389.68	1	0.037	1
	8.09	1283.7	383.4	1	0.037	1
Solution 2: Maximum kiln productivity	9.34	1395.7	379	1	0.037	1
	9.27	1315	357.5	1	0.055	1

Knowledge of the clustering of input variables and their impact on achieving a given performance function allows timely intervention by controllers to retain stability and optimal performance. The knowledge of the relationship between key system parameters (i.e. control inputs) and performance parameters (i.e. output) allows engineers/practitioners to provide alternative solutions to a given problem in a short matter of time. Furthermore, these alternative solutions, with knowledge of each solution's cost, leads the system manager to maximize profit. This unique feature is immensely important, because engineers/practitioners can select a number of good solutions, and should a solution fail to work, they will be offered alternative ones. The ability to return to stable or optimal condition in real-time using alternative solutions is unique to EventiC and may also potentially have an impact on the efficiency of control systems.

In the following section we demonstrate the application of EventiC to meet customers' satisfaction in their orders.

5.4.1. *Cement quality as a key performance indicator*

The quality of produced cement is directly related to the kiln temperature. The best cement quality is produced at a kiln temperature of 1550°C. The quality drops to medium at 1450°C and the lowest passable quality is around 1350°C.

Table 5.6 shows the clusters of system settings in which the kiln temperature average is about 1350°C. Consequently, low-quality cement is produced with these clusters of inputs. Table 5.9 shows the clusters of inputs in which kiln temperature is at the range of medium-quality cement (about 1450°C) and the EventiC algorithm optimizes the best point of set to meet the maximum production rate and minimum CO_2 emissions.

Table 5.9. Five alternative clusters of input variables for medium quality of cement

Kiln Production Rate (Tons /Hour)	Kiln Temperature (degree Celsius)	CO output (mg /m^3)	Kiln fan (on/off)	CO2 output (%)	Motors pulls material from Kiln(on/off)
8.18	1454	338	1	14.26	1
8.2	1446	340	1	14.07	1
8.36	1441	319	1	13.52	1
8.95	1452	319	1	12.52	1
8.86	1439	342	1	13.69	1

Minimum CO2 emission/ Maximum production rate

5.4.2. *Environmental impact as a key performance indicator*

Due to high demand in the construction sector the cement industry remains a major emitter of Green House Gases (GHG). The existing methods for reduction of pollutant emissions do not seem to be capable of offsetting such growth (Mahasenan *et al*., 2003). Therefore, controlling the levels of emissions is a major challenge for the industry. In this example we focus on CO_2 emissions. The key factors that determine levels of CO_2 emissions are lime production, cement kiln dust and fuel combustion patterns. The complexity is compounded in hybrid fuel systems. They may be fuelled by natural gas, coal, coke, oil, and organic material – each with a specific burning profile and emissions levels.

The amount of CO_2 emitted by the cement industry should pass the host country's regulations and international standards. Thanks to EventiC, the input cluster displayed in table 5.8shows the best system setting where CO_2 emission is at a minimum and production rate is at maximum.

5.5. The detection of unknown factors affecting the behaviour of a system

EventiC is not only an intelligent recorder of events, but is also a tool that enables preliminary data and knowledge construction. It is a complimentary middleware between the plant and its operational environmental. By including information about events that were not thought of by engineers at the outset of design and modelling, EventiC could fundamentally shift perceptions of rigid system boundaries to more dynamic boundaries.

Figure 5.11 illustrates the current approach to complex systems which breaks large systems into isolated and more abstract smaller systems for the purpose of explaining and controlling them. The principle of isolation is now becoming less practicable in modern complex systems. If a system designer is able to analyse a wider range of potential influences, then more accurate models may be able to be produced.

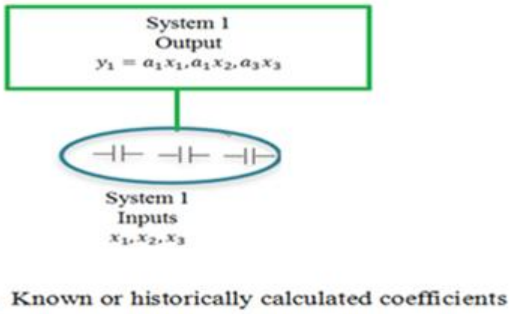


Figure 5.11. Input/output relationship in current systems

The concept of EventiC is based on managing the interrelationships and internal dynamics of the components within the eco-system of embedded systems and their environment. It achieves this automatically, thus lessening the time for detection, classification and analysis of known and previously unknown input data. Assuming that all inputs influencing the system are potentially related to the outputs of the system, this method finds potentially non-intuitive and complex relationships unlikely to be identified by conventional systems analysis. Figure 5.12 shows how EventiC could integrate isolated systems together to detect potentially unknown factors in predictive models. This feature allows engineers to build more effective, safe and responsive systems that become part of the volatile environment they function in.

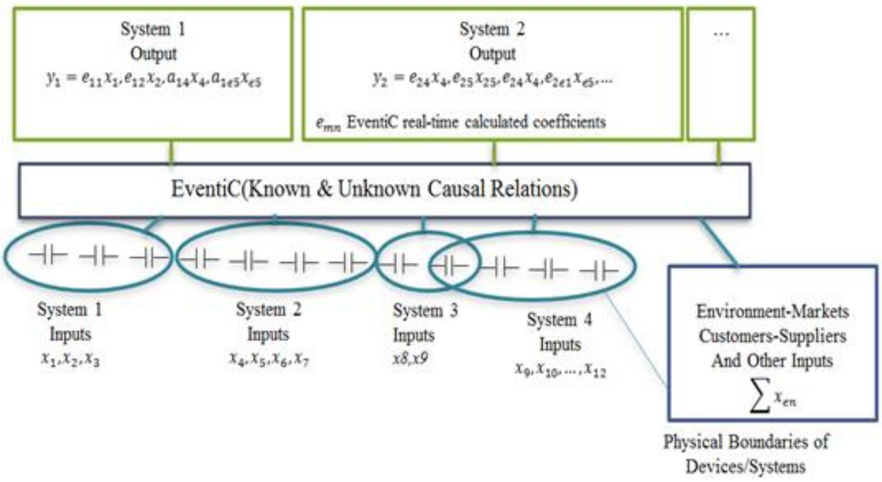


Figure 5.12. Input / Output interrelationship via EventiC

For example in our experiment, it was confirmed that the 'speed of the motor which pulls material from the kiln' has a 92% sensitivity impact on the kiln's production rate. In most cement literature, the speed of the motor's impact is not mentioned at all and is ignored. Thus one of the major advantages of EventiC methodology is recognising such influencing parameters that are unknown to controllers and modellers.

5.6. Visualisation of input/output correlation with EventiC

Big data applications need to handle many divergent types of input sources, from physical (sensor/IoT) to customer satisfaction software (ex. SAP), causing it messy, imprecise, and incomplete. Due to big data quantitative (volume and velocity) and qualitative (variety) challenges, it resembles something like “the elephant to the blind men”. It is imperative to enact a major paradigm shift in data mining and learning tools so that information from diversified sources must be integrated together to unravel information hidden in the massive and messy big data, so that, metaphorically speaking, it would let the blind men “see” the elephant (Kung, 2015).

EventiC provides a big picture to help users to visualise system's input/output correlations via decomposition of data which comes from divergent source and then composition of cause-effect clustering of system's input events (originating from sensor/actuations) and output events (i.e. performance indicators/factors). At each system's sampling scan, a matrix of I/O coincidence is produced, akin to recording a clip in a film. A time span for recording is specified, which is based on a number of observations that provides sufficient levels of confidence intervals. This visualization improves and explores the application of the proposed novel technique in a more demanding and complex environments.

Figure 5.13 illustrates this composition and decomposition.

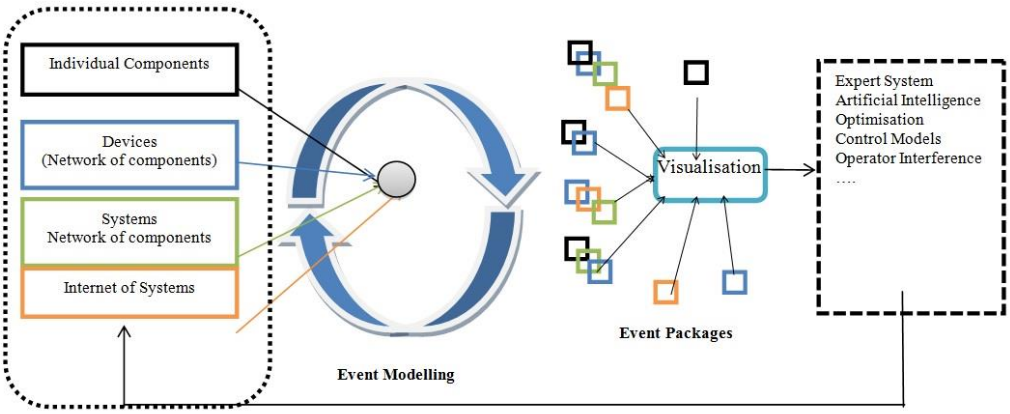


Figure 5.13. Data decomposition and composition in EventiC

5.7. EventiC application implementation

EventiC algorithm is coded in MATLAB 2012 package. This application is able to access to our industrial partner’s PLC (Programmable Logic Controllers) via industrial software called 'DataBridge'. These two software packages will be introduced shortly before presentation of the EventiC application.

5.7.1. *MATLBAB*

MATLAB (Matrix Laboratory) is a high-level language fourth-generation programming language developed by Math Works. MATLAB allows matrix manipulations and creates an interactive environment for numerical computation, visualization, and programming. With MATLAB, engineers and scientists can analyse data, develop algorithms, and create models and applications. The language, tools, and built-in math functions enable the exploration of multiple approaches and reach faster solutions than with spreadsheets or traditional programming languages such as C/C++ or Java.

MATLAB has a range of applications, including signal processing and communications, image and video processing, control systems, test and measurement, computational finance, and computational biology. More than a million engineers and scientists in industry and academia use MATLAB, the language of technical computing.

5.7.2. *DataBridge industrial environment*

DataBridge is software built by our industrial partner with a focus on the interconnection between the process control devices and monitoring /optimization systems. The software was developed for flexibility, extensibility and modularity objectives.

DataBridge combines multiple modules, ensuring a flexible and extendable working environment. Each algorithm or method is enclosed inside a plugin, also called a module, and from now onwards named a bank. The implementation and algorithm execution of each bank is independent from other banks, but the data interface is completely compatible for transmitting the data between modules and creating efficient control networks.

The DataBridge base concept is divided into group and bank modules. In each new project, DataBridge creates an acquisition group and for each group, three banks are created:

- Extract

- Transform
- Load

The cycle control execution period is defined in a group, and the modules that perform the desired control network are declared in the same place. Instead of a period control cycle, it is possible to have other types such as those triggered based on events.

The extract module or the source module, defines the type of communication between the source and DataBridge. Here, the variables that will be used in the control network are chosen, treated (transformed) by other plugins and loaded, at the end, to the output device. The software due to its modular capability and flexible permits creates different plugins to communicate with a large number of industrial communication protocols and other data servers. Examples of currently supported protocols are OPC, Modbus, Ethernet IP, CSV and even SQL.

Transform is the section used by the software to process the data from an extract module to a load module. The process algorithms are developed in these kind of modules for executing the desired tasks. For example, if you need one output that will be the sum of the two input variables, then you can use the plugin sum as a transform plug-in. The result from the sum plugin will be sent into load module, which sends the result to the desired output. The load module is the counterpart of the extract module. This provides the mechanisms to transfer the data from DataBridge to the field devices. It contains all the protocols, developed in the extract modules, such as OPC, Modbus, Ethernet IP, CSV and even SQL (DataBridge User Manual, 2011).

DataBridge’s output is stored as an excel file in MATLAB roots and is ready to feed into the EventiC application. Appendix C shows a sample of DataBridge’s output for the first 60 samples.

5.7.3. *EventiC algorithm implementation*

As explained above, EventiC application is coded in the MATLAB package. The reason for choosing MATLAB was because of its capability to manipulate matrices easily. MATLAB is also a high-level programming visual environment and DataBridge outputs were in matrix format, so the clusters’ presentation and manipulation were easily applicable for the research purpose.

Figure 5.14 illustrates the output of EventiC with respect to 196 inputs over 43,000 sampling snapshots with a 1% (0.01) sensors threshold. In the application, user set up the EventiC initiation which are: thresholds, search slot, cutting rate and input file and model's outputs or KPIs. Then the EventiC algorithm runs and presents SA weight and outputs diagrams of selected outputs and clusters of inputs/outputs with related inputs.

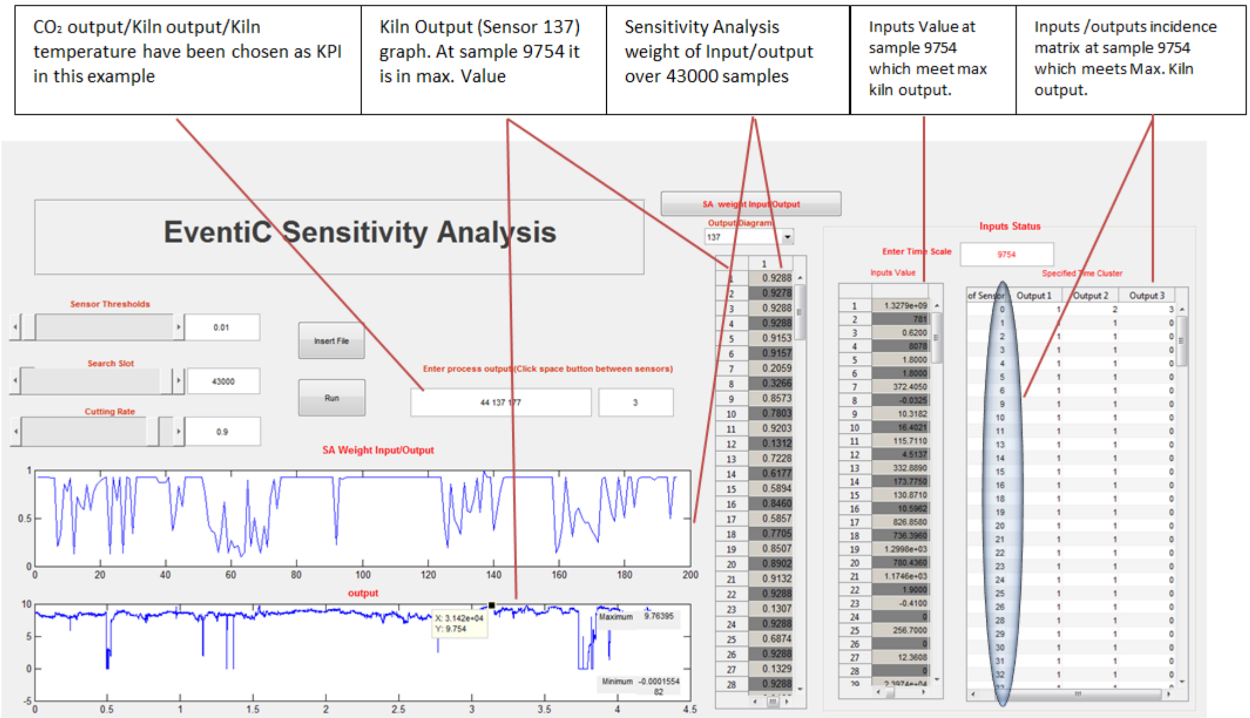


Figure 5.14. The output of EventiC with respect to 196 inputs over 43,000 sampling snapshots with 1% (0.01) sensors threshold

5.8. A Summary to the EventiC applications case study

As test beds, the application of the real-time event clustering method has been deployed and reported on a real world case study in this chapter. As an experiment, the kiln (cooking) process in the cement industry has been reviewed and the key features and ability of the real-time event clustering technique to rapidly generate an event-driven incidence matrix and measure the degrees of influence of the kiln's input sensors on various outputs has been demonstrated. The experiment equipped a kiln with 196 sensors scanned at one minute intervals. The acquired data provided EventiC with sufficient information to optimize the number of relevant input variables and provide accurate knowledge of the system state over a 1 month period (43,000 observations). The results of the EventiC algorithm shows that 18% of TDs have little effect on the kiln production rate, therefore they can be totally ignored when measuring the kiln's production rate. This experiment has been repeated with regard to the kiln's energy consumption and CO_2 emissions as a major environmental pollutant.

EventiC's advantages over plant control and production optimization has been reviewed and the following three main benefits can be summarised.

1. A minimization of the negative consequences of variations in energy sources and raw materials by improving the production process and quality monitoring, control, and specialized optimization methods.
2. A reduction in energy consumption and an increase in resource utilization by deploying new processes of monitoring, control and optimization techniques.
3. A reduction in the industry's environmental impact through clever usage of raw materials and energy.

EventiC achieves this by: (a) interpreting changes in the values of input-output (I/O) data at the event level, (b) detecting if I/O events coincide, and (c) groups I/O events as related events. This processing happens in a specified time interval, known as the scan rate whose duration can potentially range from microseconds through to minutes. At each scan a matrix of I/O coincidence is produced, akin to recording a clip in a film scenario. This scenario is unique and could be assumed as an ID for each I/O coincidence a time span for recording is specified, which is based on a number of observations that provides sufficient levels of confidence intervals (e.g. 250 clips). The weighting of an input on output is calculated as the number of coincidences in the time span. Once the relationship between the inputs, outputs and their weightings are established, for the purpose of modelling and control we revert back to the actual value of the inputs and the outputs. The translation of system parameters to events and grouping the relevant I/O events in near real-time may be considered as a novel approach in the understanding and processing of large-scale raw data/signals.

In the next chapter the application of the proposed technology will be used as an automatic input variable selection tool in fuzzy controllers.

6. An Application of EventiC in Input Variable Selection: A case study for Fuzzy Controller

The purpose of this chapter is to integrate the EventiC algorithm to a Fuzzy Logic Controller (FLC) designed for rotary cement kilns (based on the real behaviour of a cement kiln) to automatically extract all fuzzy parameters and reduce the rule base of fuzzy control complex systems. The FLC learning is performed by the proposed EventiC through a set of controlled input/output data. Causal relationship modelling and parameter weighing mechanism of the proposed EventiC method has been used to extract fuzzy control inference rules. Therefore, EventiC could be utilised as a more cost-effective alternative for the input variable selection of fuzzy controllers.

A conventional controller needs a mathematical model, which is either very difficult or frequently impossible to obtain. Whilst a relatively accurate model of a dynamic system can be developed it is often too complex to adopt in a controller development, especially on the many control design procedures that require restrictive assumptions for the plant. There are two main reasons that hamper the design and development of successful model-based controllers in industry. Firstly, the complexity of interpreting the system state may not necessarily lend itself to direct model-based control solutions, as at times these solutions prove to be unrealistic and impractical to implement in the real world. Secondly, useful heuristics are sometimes ignored by modellers because they do not fit into a specified mathematical framework. To overcome these shortcomings the industrial fuzzy controller as an expert controller (using semantic rules, based on the fuzzy control theory) can be adopted. This provides a formal methodology for representing, manipulating and introducing human knowledge about how to control a system without heavy reliance on complex mathematical modelling. Therefore one can view the fuzzy controller as an artificial decision maker that operates in a closed-loop system in real-time. To design the fuzzy controller, the control engineers must gather information as to how the artificial decision maker should act in the closed-loop system. In manufacturing environments this information derives from a human operator who performs the control tasks. This said, on occasion the control engineer can come to an understanding of the plant's dynamics and write down a set of rules about how to control the system without outside help. However, the complexity of kiln operations does not lend itself to classical control functions. As an alternative the industry has adopted fuzzy controllers. For the most part fuzzy controller inference rules are derived either from direct expert knowledge of the process, or automatically by techniques such as neural networking or genetic algorithms. These methods generally rely either on historical data derived by experience or machine-learning algorithms such as in the case of neural networks and genetic programming.

Thanks to the proposed Event Clustering method, its causal relationship modelling and parameter weighing mechanism, bodes well with fuzzy control inference rules. Therefore, EventiC sitting top of a typical data acquisition system (e.g. SCADA) is fully capable of automatically translating into cause-effect events. The combination of these event sets (EventiC inputs) forms a set of performance variables (EventiC outputs) and the fuzzy inference rules and parameters will be made available to the fuzzy modeller. As an alternative to its own simple optimizer EventiC can also function as an improved IVS method to both classical and modern (e.g. numerical, analytical, genetic, fuzzy, etc.) controllers-optimizers. Not exclusive of fuzzy controllers, we will describe how EventiC can be used to extract the parameters of a FLC.

This chapter will be structured by an initial summary of fuzzy controllers and their application in the cement production process. Following this, the issues and variables to be defined and controlled in a rotary kiln are addressed. Finally, in the latter section of the chapter, an overview of the proposed EventiC-fuzzy controller (EFC) for the rotary cement kiln is presented.

6.1. Fuzzy systems

Fuzzy theory was introduced in 1965 by Lotfi A. Zadeh in his paper 'Fuzzy Set' (Zadeh, 1965).Zadeh proposed the fuzzy algorithms concept, fuzzy decision making and fuzzy ordering respectively in 1968, 1970 and 1971. In 1973, Zadeh proposed the foundation for fuzzy control in his paper, 'Outline of a new approach to the analysis of complex systems and decision process' (Zadeh, 1973)

Mamdani and Assilian (Mamdani&Assilian, 1975) applied the basic framework of a fuzzy controller to control a steam engine. In later decades, several studies have been presented that show the application of fuzzy controllers to control complex nonlinear processes that cannot be easily modelled by mathematical equations. Recently, FLCs have been used for a wide variety of industrial system and consumer products, from control to signal processing, communications, manufacturing and expert systems for business. However, the most significant applications have concentrated on control. FLCs are rule-based systems which can be useful to control those systems which are controlled by skilled human operators without any mathematical knowledge of the process dynamic (Herrera et al., 1995).

FLCs are based on a set of fuzzy control rules that draw upon an expert's common sense and experiences. However, there still exist many difficulties in designing fuzzy systems to solve certain complex nonlinear problems. For a better understanding, an outline of the limitations of conventional controllers and the benefits of fuzzy controllers are provided in Table 6.1.

Table 6.1. Limitations of conventional controllers versus the benefits of fuzzy controllers

Limitations of conventional controllers	Benefit of fuzzy controllers
• Plant nonlinearity. The efficient linear models of the process or the object under control are too restrictive. Nonlinear models are computationally intensive and have complex stability problems. • Plant uncertainty. A plant does not have accurate models due to uncertainty and lack of perfect knowledge. • Multivariable, multi-loops and environment constraints. Multivariate and multi-loop systems have complex constraints and dependencies. • Uncertainty in measurements. Uncertain measurements do not necessarily have stochastic noise models. • Temporal behaviour. Plants, controllers, environments and their constraints vary with time. Moreover, time delays are difficult to model.	• Fuzzy controllers are more robust than PID controllers because they can cover a much wider range of operating conditions than PID can, and can operate with noise and disturbances of different natures. • Developing a fuzzy controller is cheaper than developing a model-based or other controller to do the same thing. • Fuzzy controllers are customisable, since it is easier to understand and modify their rules, which not only use human operator's strategy but also are expressed I natural linguistic terms. • It is easy to learn how fuzzy controllers operate and how to design and apply them to a concrete application. • Smooth controller behaviours • Robust controller behaviours • Ability to control unstable systems

6.1.1. *Fuzzy system concept*

Fuzzy logic is a form of many-valued logic and deals with reasoning rather than fixed and exact logic. Compared to traditional binary sets, fuzzy logic variables may have a value that ranges in degree between 0 and 1. Figure 6.1 shows the range of logical values in Boolean and fuzzy logic where the value may range between completely white and completely black.

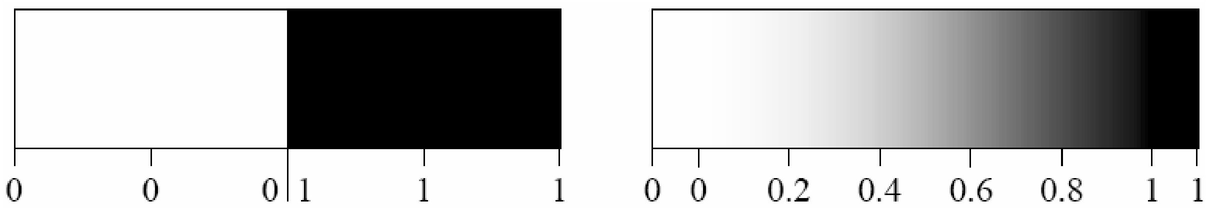


Figure 6.1. Range of logical values in Boolean (left) and fuzzy logic (right)

Fuzzy Logic Systems (FLS) are knowledge-based systems which use an expert's common sense and experiences in the form of fuzzy IF-THEN rules. These systems are typically characterized by a group of four main elements:

- Knowledge-Base
- Fuzzifier
- Inference engine
- Defuzzifier

This general scheme can be seen in figure 6.2.

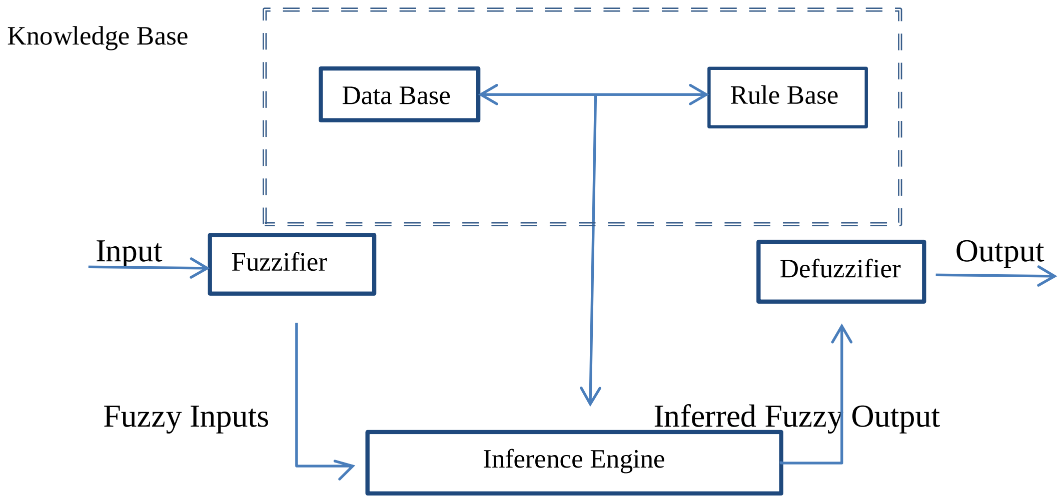


Figure 6.2. General scheme of a Fuzzy Controller (FC)

In the following sections, a brief explanation of the elements that constitute the FLS will be performed.

6.1.1.1 Knowledge-Base

The knowledge-base is one of the most important components of a fuzzy system, since all other components rely on it. The fuzzy rules are composed in two parts: the antecedent (IF part) and the consequent (THEN part).Therefore, a knowledge-base composed by a set of N fuzzy IF-THEN rules R:

$$R_j\text{: IF } x_1 \text{ is } A_{j1}, \ldots \text{ and } x_n \text{ is } A_{jn} \text{ THENU is } B_j \qquad \textbf{Equation 6.1}$$

Where A_{ji} and $B_j$are the linguistic terms characterized by fuzzy membership function $\mu_A$in (x) =U---> [0, 1] and μBj(u)=U---> [0,1] , respectively j=1,2,…,N ,where N is the number of fuzzy rules; xi(i=1,2,..,n) are the fuzzy system input variables, and U is the output.

The most commonly used membership function types are the trapezoidal, triangular and Gaussian membership functions, as represented in figure 6.3.

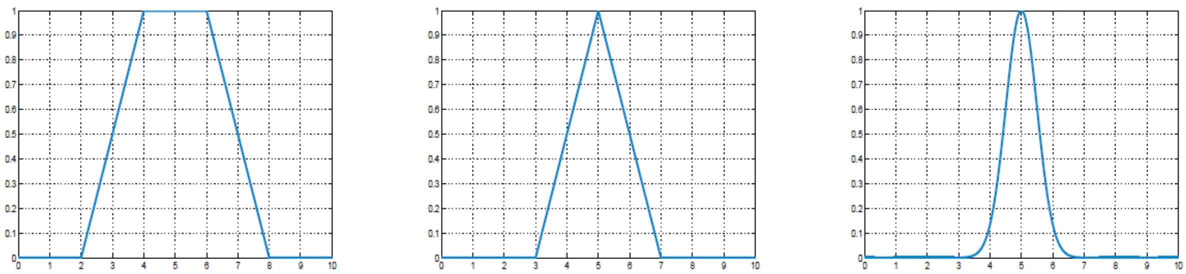


Figure 6.3. Examples of membership functions: a) Trapezoidal b) Triangular c) Gaussian

Reznik (*et al*., 2000) classified the designing of the fuzzy logic rule base as needing to consist of 4 procedures:

1. Determining the process states and control variables of the system.
2. Determining the right input variables for the controller.
3. Creating a suitable fuzzy logic rule for the specific system (IF-THEN rule).
4. Creating the fuzzy interference engine (this will be explained in detail, later in this section).

6.1.1.2 Fuzzifier

The next fuzzy system element is the fuzzifier. This element is responsible for mapping the real values of the input linguistic variables, x, into corresponding fuzzy sets described by membership functions X. The fuzzifier has the main goal of transforming the input real value

Ř in a fuzzy set as defined by a universe of discourse S. There are many methods to fuzzify the inputs, sometimes as in equation 6.2, only the singleton fuzzifier is considered, due to its simplicity of implementation. However, other fuzzifier methods can be consulted in (Reznik, 1997)

$$\text{Singleton fuzzifier: } \mu \text{ A0(x)} = \begin{cases} 1, & if\ x = x \\ 0, & other\ cases \end{cases}$$

Equation 6.2

Where x is the concrete input value.

Another way which is also comparatively simple for implementation is the Gaussian Fuzzify method which has been used in much research due to its simplicity of the statistical input data (Wang & Mendel, 1992). Figure 6.4 shows the formula of Gaussian membership functions and its corresponding graph, where w_c is the mean or centre value of the value, and wd is the reciprocal of the standard deviation of the function (w_s is the connection weight, which in normal situations is equal to 1) (Watanabe *et al.*, 1996).

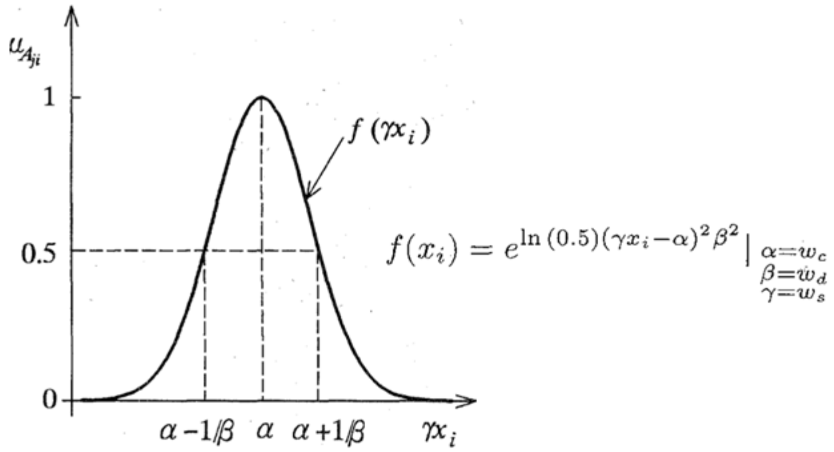


Figure 6.4. Gaussian Membership function

The output of this element will be the input of the interference engine part which requires the fuzzy subset inputs to be able to function with the fuzzy logic rules of the FCS.

6.1.1.3 Fuzzy Inference Engine (FIE)

The FIE uses the collection of fuzzy IF-THEN rules to map the fuzzy input set X into the fuzzy rule consequent Bj. The collection of the fuzzy rule outputs are then combined into an overall inferred fuzzy output U. Generally, each rule will be processed individually according to the state of input creating. The basic methods to process the antecedent part (IF part) are: intersection, union, and complement, which are defined by t-norm, s-norm and c-norm, respectively. Table 6.2 shows the most common operator, the t-norm operator, for the Intersection method.

Table 6.2. Formula used in t-norm (intersection)

t-norm (intersection)	
Minimum	$\min(\mu_{A1}(x_1), \mu_{A2}(x_2))$
Bounded Product	$\max(0, \mu_{A1}(x_1) + \mu_{A2}(x_2) - 1)$
Algebraic Product	$\mu_{A1}(x_1)\mu_{A2}(x_2)$

After the antecedent value calculation, the implication operator will be implemented to interpret the fuzzy propositions as the fuzzy relation. The methods which are commonly used for the implication operator are Mamdani implications as shown in table 6.3.

Table 6.3. Formula used in implication operator

Implication	
Minimum	$\min(\mu_{A1}(x_1), \mu_{A2}(x_2))$
Product	$\mu_{A1}(x_1)\mu_{A2}(x_2)$

Next, in order to produce an overall output fuzzy set of the system, the result of all fuzzy rules will be aggregated using an aggregation operator. The output will be used as an input of the defuzzifier in the next section. The methods to be implemented for aggregation are presented in table 6.4.

Table 6.4. Aggregation formula

	Aggregation
Bounded Sum	$\min(\mu_{A1}(x_1) + \mu_{A2}(x_2), 1)$
Maximum	$\max(\mu_{A1}(x_1), \mu_{A2}(x_2))$
Normalized Sum	$\frac{\mu_{A1}(x_1) + \mu_{A2}(x_2)}{\min(\mu_{A1}(x_1) + \mu_{A2}(x_2), 1)}$

6.1.1.4 Defuzzifier

Finally, the defuzzifier is the fuzzy system element responsible for mapping the output fuzzy set of the FIE into a real value output. It calculates all the results from the interference engine according to the input and combines them to produce the most feasible result. The result of this method will be in the form of the crisp value for system adjustment.

The method of defuzzification is different according to the context of the system. One of the most popular approaches to defuzzification, is the implied fuzzy set, Centre of Gravity (COG) and this is used in many designs of fuzzy controller. Equation6.3 shows the classic way to calculate the crisp output by this method; moreover, it doesn't require much mathematical effort.

$$u_{Crisp} = \frac{\sum_i x_i \int \mu_i}{\sum_i \int \mu_i} \qquad \textbf{Equation 6.3}$$

6.2. Fuzzy system controller

Fuzzy controller is the controller operating on the knowledge-base, utilizing the fuzzy logic in order to convert the expert knowledge and information into an automatic control system. Some of the most important reasons regarding the attraction of fuzzy controller implementation are its high capability in operating within different system environments, the low cost of developing the controller, rules which are easy to understand and modify whilst still respecting the nature of the system, and the general ease of the concept by which the non-specialist can design, operate, and implement it.

6.2.1. *Types of Fuzzy Controller*

Reznik (1997) classified fuzzy controllers into three types according to their features;

A. Simple Fuzzy Controller

A single-level structured controller with no hierarchical rules would be counted as a simple fuzzy controller. Its lack of adaptive ability means that most of the features of this type of controller are fixed, such as the input and output scaling factor, and even the input and output themselves. The numbers of the input class are fixed and influence on the numbers of the controlling rules.

B. Complex and /or Multilevel Fuzzy Controller

Complex controller features are mostly similar to the simple controller. The only difference is that this type of controller will have a more complicated structure than the simple one. It possesses a hierarchical rule structure and multilevel structure with few controllers within.

C. Adaptive and/or self-organizing Fuzzy Controller

With an adaptive ability, this type of controller can change according to the actual system. Most of the features are theoretically flexible and can be tuned to be more compatible with the system, for example, the shape of the membership function of each class. However, there are still some limitation; practically, the input and output of the controller still need to be fixed.

6.2.2. *Design of membership function*

Regardless of the differences between each type of controller, a crucial step in the design of fuzzy rule based systems is to derive the desired fuzzy rule base. The knowledge required for the rule base can be obtained either from human experts or from measuring instruments in the form of numerical data. Therefore, there are two possible ways for designing fuzzy rule based systems.

6.2.2.1. Subjective approach to fuzzy rules generation

In this approach, the source for deriving the linguistic rules is direct expert knowledge and intuition about the system. It is this knowledge that is expressed in the form of logical IF-THEN rules. The fuzzy controller developed by Mamdani in 1975 (Mamdani&Assilian, 1975), modelling the performance of a human operator, is considered to be the first example of this approach. Since then, various fuzzy modelling techniques for the design of controllers have been developed (Sugeno, 1985; Yager & Filev 1994).

The advantage of this approach is that it is intuitive and natural. Also, the rules are more interpretable by other users because they are written in natural language, so any modifications, if required, can easily be done.

6.2.2.2. Objective approach to fuzzy rules generation

In general, it is not easy to determine the most suitable fuzzy rules and membership functions to control the output of a plant, when the only available knowledge concerning the process is the empirical information transmitted by the human operator. During the last two decades, rule generation or knowledge extraction from numerical examples has been intensively developed by various researchers (Shi &Mizumoto, 2001; Rojas et al., 2000; Aliev *et al*., 2001; Delgado *et al*., 2009; Takagi &Sugeno, 1985). Most of these methods have involved iterative learning procedures or complicated rule generation mechanisms such as the neuro fuzzy learning method, genetic-algorithm based methods, and the fuzzy c-means method (Krishnapuram & Freg, 1992).

Neuro-fuzzy learning algorithms derived from the Artificial Neural Network (ANN) algorithm to determine its parameters (fuzzy set and fuzzy rules) by mimicking the actions of an expert who solves complex problems, in other words, any ANN should be trained. Training is performed with a presentation of the sampling data. In some cases, thousands of data examples may be required to be presented in a randomised order, and the learning techniques often demand human supervision to guarantee convergence (Wang & Li, 2003; (Mitra & Hayashi, 2000). This automates also speeds up the knowledge acquisition process. Such models help in minimizing human interaction and the associated inherent bias during the phase of knowledge base formation and also reduce the possibility of generating contradictory rules. This technique can adaptively adjust membership functions and fine-tune rules to achieve better performance. The disadvantages of this approach are that ANN may require a high computational power and a long period for training, both of which are not available in some control applications. There is thus a need to optimize the training process, and the application of genetic algorithms (GA) is one of these ways.

The combination of fuzzy logic and genetic algorithms (Aliev et al., 2001; Barajas & Reyes, 2005; Juang & Lu, 2005; Mendes et al., 2011: Herrera et al., 1995) allows an optimal number of fuzzy rules in rule base and optimal values for centres and shapes of membership functions. Another advantage of genetic algorithms is that they do not require the differentiable function for optimization. The auto extraction of a fuzzy control system for industrial processes as proposed in (Mendes et al., 2011) uses GA to manipulate the parameter selection of the fuzzy system. It consists of a 5-level-hierarchical structure model. The first layer represents the number of input sets and their time delay. The second level matches the inputs to their antecedent and consequent fuzzy membership functions according to the fuzzy rules. The third and fourth levels are responsible for the individual rule and rules set selection regarding the input population from the first level. Finally, the fifth level represents the information which is necessary for the fuzzy controller, such as the aggregation method of the antecedent, the interference engine, the defuzzifying method, and the data from the previous level. Another method for the auto extraction rules of the fuzzy control systems is introduced in (Carmona *et al.*, 2010) using evolutionary algorithms.

The disadvantages of these methods are the slow convergence and longer learning times of genetic algorithms because of the large number of rules to proceed. In this case, all rules are built using all possible combinations of fuzzy input values.

Fuzzy clustering is considered to be one of the most important techniques for the automatic generation of fuzzy rules from numerical examples (Hoppner&Klawonn, 2000). This algorithm forms a fuzzy partition of data points into a given number of clusters. Each cluster represents one rule of the rule base. The number of cluster centres is the number of rules in the fuzzy system. In this way, the rule base size can be easily controlled through the control of the number of cluster centres. The disadvantage is that no clustering algorithm provides the means to determine the number of clusters, and hence the number of rules, in a fuzzy model.

The above mentioned automated techniques for the generation of rules from numerical data are very efficient provided the available data is sufficient enough for training the model. However, in many environmental problems specifically related to noise pollution, the available data sets are very limited. An attempt has been made in (Zaheeruddin & Anwer, 2005) to develop a fuzzy model for such real situations. This technique can be used as an alternative to develop a model when available data may not be sufficient to train the model.

6.2.3. *Fuzzy rule reduction*

A fuzzy controller's knowledge-base includes all the possible combination rules of fuzzy input values. Inevitably, the size of the rule grows exponentially as the number of controller inputs grows. As the complexity of a system increases, it becomes more difficult and eventually impossible to make a precise statement about its behaviour. As an example, consider a fuzzy controller, where number of inputs (n) =6 and linguistic values (m) =4. The total number of fuzzy rules will be $K = m^n$ $=4^6$=40964. If we have 5 inputs the K=1024, i.e. 75% reduction and with 4 inputs k=256, we will save 93.8%.

One of the simplest and oldest ways to reduce the rule size is the issuing of the Sliding Mode Control (Glower &Munighan, 1997). This approach has its disadvantages as the parameters for the switch function have to be selected by an expert, or designed through classical control theory (Hung et al., 1993).

Jamshidi (1997) proposed the use of sensor fusion to reduce a rule base size with a combination of several inputs into one single input. The rule base size is reduced because the number of inputs are reduced. He also proposed to use the combination of hierarchical and sensory fusion methods. The disadvantage of the design of a fusion hierarchical fuzzy controller is that much reliance has to be put on the experience of the operator to establish the needed parameters.

Ledeneva (2006) proposed three methods for finding the combination of the parameters for the rule base reduction methods. He automated estimation of sensory fusion and hierarchical methods and developed new areas of control to work more efficiently. His method offers great rewards to model nonlinearities without the necessity of using complex algorithms.

6.3. Fuzzy controller in the cement industry

The basic process in a cement production plant is the baking of the raw material mix in a kiln. Cement kilns exhibit time-varying nonlinear behaviour in which both the physical and chemical reactions occurring in the kiln complicate its dynamic equations. The corresponding equations have not been derived completely and accurately, while a lot of present variables are discarded in the equations (Noshirvani *et al*., 2009; Fallahpour *et al*., 2008). It is obvious that this may cause various problems in designing the controller for rotary kilns.

Although cement is the final product of a cement factory, the output product of the kiln is called clinker. Producing high-quality clinker, realising efficiency improvement in input material and energy consumption are the goals of the cement industry. These are all achievable by deploying a desirable and appropriate kiln controller. Producing neither over burning nor under-burning clinker is acceptable for a rotary kiln. Some factors such as flame shape, secondary air temperature, ID fan speed all have considerable effects on the clinker quality (Fallahpour *et al.,* 2008).

Most cement factories throughout the world are controlled by the direct knowledge of expert kiln operators. Therefore, having accurate knowledge of the situation and state of the burning zone is critical for a kiln operator.

The use of fuzzy controllers in cement kilns was one of the first successful applications of the fuzzy controller in industry. In 1978, Holmblad and Ostergaard used the first fuzzy controller for a complex industry process: the cement kiln. They saw that the results were much better than when the kiln was directly controlled by humans (Wang, 1994).

Nowadays, the case for using fuzzy logic controllers to control cement kilns has increased. This is based on the fact that fuzzy logic controllers do not need an accurate model of the plant. By using a fuzzy logic controller, a remarkable improvement in cement quality and a decline in production expenses has been achieved. Several designs of such controllers have been proposed and/or implemented over recent years, which have been designed and based on the knowledge of the operators. Image processing has been proposed as a solution to control cement kilns.

6.3.1. *Cement production process*

Cement production process is comprehensively explained in chapter five. A short review of the process from the perspective of cement quality will be explained in this section. Figure 6.5 shows the structure of a rotary kiln with the most important variables used for control purposes in simplified form (Fallahpour et al., 2007). The kiln is a long and complex tunnel, generally with a cylindrical shape. The input materials include carbonates and silicates which should be burned to generate solid oxides and combustion gases. The burning process, which denotes to all activities done on the raw materials up to making the final clinker, is done in the pre-heater, kiln and cooler (Noshirvani et al., 2009). Raw meal must be preheated and completely dried before it is fed into the kiln. Hot smoke generated in the kiln during the clinker production is used to do that. The pre-heater is responsible to acquire the remaining moisture of the raw material and break up silicates, as well as partially calcinating the present carbonates in the material.

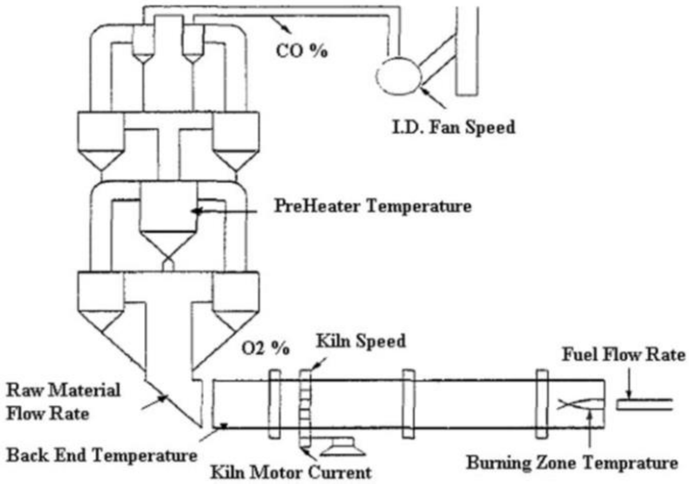


Figure 6.5. A rotary kiln plant

The main part of the burning is completed in the kiln. The kiln has two baking furnaces, which are the back-end and the burning zone. The calcification of raw material is first completed in the back-end and then fed into the main baking furnace.

The cement kiln is a large cylindrical tunnel with its size is directly related to the factory size. The cylinder with a slope of about 4%, rotates around its axis and the raw meal dust sticks to its walls. This is then gradually baked and transformed to clinker (which is grained pieces of cement). This is transported away from the kiln and milled in a special mill to produce cement dust. Finally, the cylinder is slightly inclined down with clinker and cement dust slipping towards the cylinder output. On the side of the kiln, a flame is set to heat up the kiln between 1350°C-1550°C and tiny air tunnels control the oxygen content in the kiln. On the other side of the kiln, suction fans are set up close to the pre-heater cause entering the air. This affects upon the flame direction and gas combustion.

Calcification is the first chemical reaction done on the raw mill at the kiln's high temperature. The high temperature at the burning zone melts the input raw materials. Then, the main burning is gradually started and chemical reactions occur between the silicates and the present oxygen of the air. CO gas includes the main part of the combustion smokes. Finally, the cement crystals are made and go out from the kiln as the clinker (Fallahpour*et al*., 2007).

$$CaCO_3 + heat \rightarrow CaO(s) + CO_2\ (g)$$ **Equation 6.4**

The output clinker has a temperature from 1000-1200°C and is cooled at the end of the kiln ready to be transferred. Cooling the clinker has an effect on its quality. The main goal of the kiln control is to produce high-quality clinker. On the other hand, perfect control of the burning zone temperature is also an important factor regarding the quality of clinker. The control variables will be discussed in the next section.

6.3.2. *Select effective input and output variables in the cement quality control*

In this thesis, a black box identification procedure for modelling the kiln production has been used. A review of literature and expert knowledge in the cement production process confirms five secondary variables (named KPI and explained in chapter five) which are controlled by nine other variables, effect on cement quality. Figure 6.6 classifies these variables. However, it should be noted that in cement kiln systems, the dependency degrees of inputs to outputs are not the same for all variables.

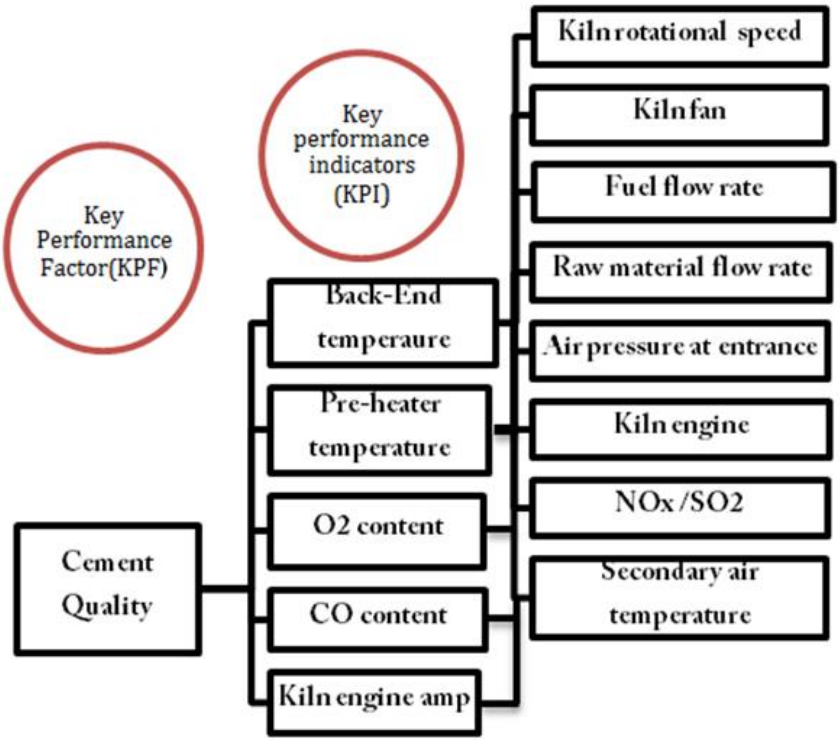


Figure 6.6. Cement quality key performance indicators and relevant inputs

Figure 6.6 shows that cement quality will be under controlled if nine control variables are controlled. These nine control variable are:

- Raw material feed rate (MAT) which represents the raw material feed to the kiln.
- Kiln fuel (FUEL) which is the fuel consumed to heat up the material in the kiln.
- Rotational speed (KS) which controls the speed of material in the kiln and adjusts the volume of the input materials to the rotary kiln and the feed rate of the raw mix. In the case of more input materials, the kiln speed should be adjusted to complete the burning process.
- Kiln fan (FAN) which controls the flow of fresh air into the kiln and adjusts the required O_2 percent of burning in the system. Furthermore, blowing the air into the kiln causes CO gas out from the kiln and consequently, the reduction of the CO density in the kiln.
- Kiln engine pulls the material output (KM) which is an engine in the kiln which pulls the material to output.
- The secondary air temperature (AT) which is another input for the rotary kiln which blows air into the kiln. This hot air is usually provided during the cooling of hot clinker.

- The secondary air pressure (AP) which causes an increase in the present oxygen volume via the parameter of O_2 in the kiln.
- SO_2 which are injected to process to improve the cement quality
- NOx which are injected to process to improve the cement quality

A summary of these kiln inputs including variable names which represent the variables in the rest of thesis are shown in table 6.5.

Table 6.5. Kiln inputs from cement quality perspective

Kiln Input (control) variables	Variable Name
Kiln Rotational speed(rpm)	KRS
Kiln fan (On/Off)	FAN
Kiln Fuel (Ton/hrs.)	FUEL
Raw material feed rate (%)	MAT
Secondary air pressure	AP
Kiln engine pull the material output (%)	KM
Secondary air temperature	AT
Injected NOx to kiln	NOx
Injected SO2 to kiln	SO2

The kiln operation produces five outputs to control the clinker with accepted quality, which are:

- The % of CO at the pre-heather (CO) which represents the combustion efficiency.
- The amperage of the kiln engine (KA) which indicates the current at which the kiln consumes, is proportional to the amount of both the kiln's speed and load.
- The pre-heater temperature (PT) which represents the temperature at the pre-heater.
- The back-end temperature (KT) which is related to the quality of the produced clinker.
- Oxygen gas content (O_2) which is the dependence same as CO gas. This depends on ID fan speed, fuel and secondary air pressure in the kiln, as well.

A summary of these kiln outputs including variables names, are shown in table 6.6.

Table 6.6. Output variables of the kiln

Kiln outputs	Variable Name
Back-end Temperature	KT
Pre-heater temperature	PT
O2 content in the kiln	O2
CO content at pre-heater	CO
Kiln Engine motor amp	KA

6.4. EventiC integration to cement fuzzy controllers

As explained previously, the fuzzy controller's knowledge-base includes all the possible combination rules of fuzzy input values with the size of the rule depending exponentially on the number of controller inputs. As the complexity of a system increases, it becomes more difficult and eventually impossible to make a precise statement about its behaviour. It is reasonable to consider though that the inputs of controllers have an effect upon it. Following on from this, the inputs which have either little influence or no influence on outputs should be discarded.

6.4.1. *Input variables delay*

The cement kiln has delays in its operation (Makaremi et al., 2008). In other words, the results and effects of input variations may appear in the outputs with different delays. To have a controller with a high ability in the control of the real system, delays should be considered. However, inputs have different effects on each output, and so have different delays. Makaremi et al. (2008) used Lipschitz quotients to find the input delays in cement production. The results are shown in table 6.7.

Table 6.7. Input variables delay estimation

	Delays (Min)			
Variable	BT	PT	CO& O_2	KA
KRS	36	40	5	0
FAN	0	5	0	0
FUEL	4	10	5	25
MAT	18	30	15	25
AP	0	5	3	30
KM	0	0	0	0
SO_2/NOx	0	0	0	0
AT	0	0	0	0

6.4.2. EventiC application

The 196 sensors and actuators that provided the raw data input received from our industrial partner's SCADA has been fed into the EventiC application to find the effects of table 6.5's input weights on table 6.6's output variables in the cement process with relevant delays, as mentioned in table 6.7. The results of the EventiC output are shown in table 6.8.

Table 6.8. Categorization of output and input variables

Sensor Type	BT	PT	O_2	CO	KA
KRS	**0.92**	**0.28**	**0.56**	**0.49**	**0.91**
FAN	**0.93**	**0.92**	**0.98**	**0.90**	**0.41**
FUEL	**0.95**	**0.96**	**0.93**	**0.91**	**0.31**
MAT	**0.92**	**0.25**	**0.27**	**0.25**	**0.90**
AP	**0.91**	**0.46**	**0.98**	**0.95**	**0.61**
KM	0.24	0.65	0.52	0.32	0.42
IH	**0.97**	**0.91**	**0.47**	**0.32**	**0.84**
AT	**0.96**	**0.92**	**0.55**	**0.44**	**0.36**
NOx	0.45	0.46	0.46	0.46	0.53
SO_2	0.27	0.25	0.25	0.25	0.37

As explained in chapter five, the proposed method challenges expert decisions. In this experiment, as shown in table 6.9, with CT=90%, three inputs (KM, NOx and SO_2) have less effect on the outputs and could be discarded from the cement quality model. However, "I/h return in the kiln (IH)" which has not been considered in the literature, demonstrates an impact on the kiln's output quality. Reducing 2 out of 10 input variables, assuming 3 linguistic variables for the fuzzy controllers, decreases the fuzzy rule $\frac{3^9}{3^7} = 9$ times and hence the controller synthesis 9 times faster.

Table 6.9. Categorization of the kiln's input and output variables with CT=90%

Sensor Type	BT	PT	O_2	CO	KA
KRS	*				*
FAN	*	*	*	*	
FUEL	*	*	*	*	
MAT	*				*
AP	*		*	*	
KM					
IH	*	*		*	*
AT	*	*			
NOx					
SO_2					

6.4.3. EventiC/fuzzy controller structure

The structure of the proposed integrated EventiC/fuzzy controller system for the cement kiln is shown in figure 6.7. Designing the intelligent controller of the kiln has been based on fuzzy logic. The design requires an identified model of the plant which has a perfect representation of the real characteristics of the kiln. It is noted that based on the EventiC output, the model has a five multi input single output (MISO) system which has seven inputs and one output.

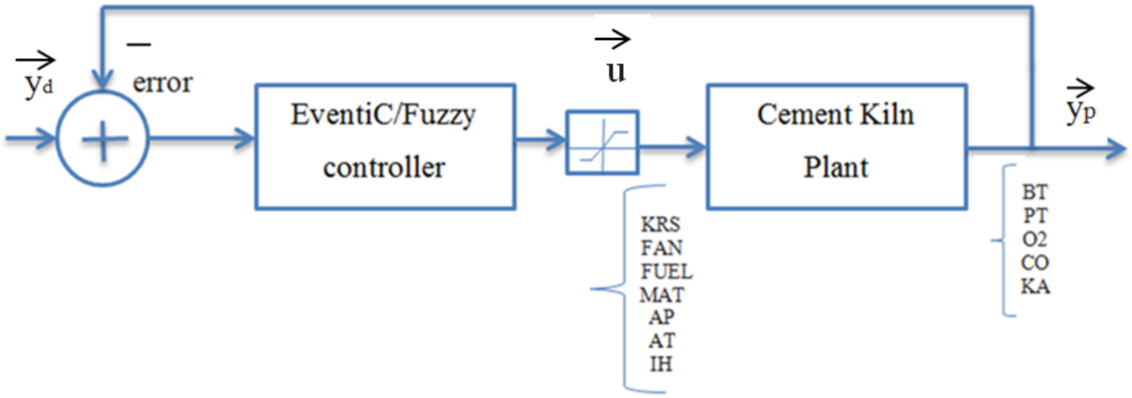


Figure 6.7. EventiC/fuzzy controller integrated structure

As figure 6.7 shows the control scheme for a kiln plant, the set-points variables yd are the reference signals and applied to the input of the system. These are what the controller tries to track and reach. The vector u is the inputs mentioned in table 6.6 and denoted as the output of the controller. Finally, the vector y_p includes the output variables of the plant which have the normal variables introduced in table 6.5.

The categorization of input/output presented in table 6.9 could be segregated in tables 6.10, 6.11 and 6.12.

Table 6.10. Kiln's inputs relationships with O_2 and CO

Sensor Type	O_2	CO
FAN	*	*
FUEL	*	*
AP	*	*

Table 6.11. Kiln's inputs relationships with KA

Sensor Type	KA
KRS	*
MAT	*
IH	*

Table 6.12. Kiln's inputs relationships with BT and PT

Sensor Type	BT	PT
KRS	*	
FAN	*	*
FUEL	*	*
MAT	*	
AP	*	
IH	*	*
AT	*	*

The controller designed in figure 6.7 based on the above description can be simplified and segregated into three distinct EventiC/fuzzy controllers (as presented in figure 6.8) which can then be used to control different variables of the kiln. In the first controller, the control effort signals are generated from the error signal of PT and BT. In the second and third controllers, the measured values for CO and $O_2$and KA are compared with their authorized values. If the values are not in the normal range, the corresponding controller controls and generates the required control signals. These signals are the average of the outputs of the three EventiC-fuzzy controllers depicted in figure 6.8.

Literature reviews in cement quality show that there are no reference set-points for O_2, CO and the electrical current of the kiln (KA), since reaching a specific reference input for these variables at the controller is not a target for the control scheme. However, there are normal authorized ranges that passing may cause problems in cement production. If the value is not in the normal range, the corresponding controller acts and generates the required control signals. The ranges have been listed in table 6.13 (Ghosh, 2003).

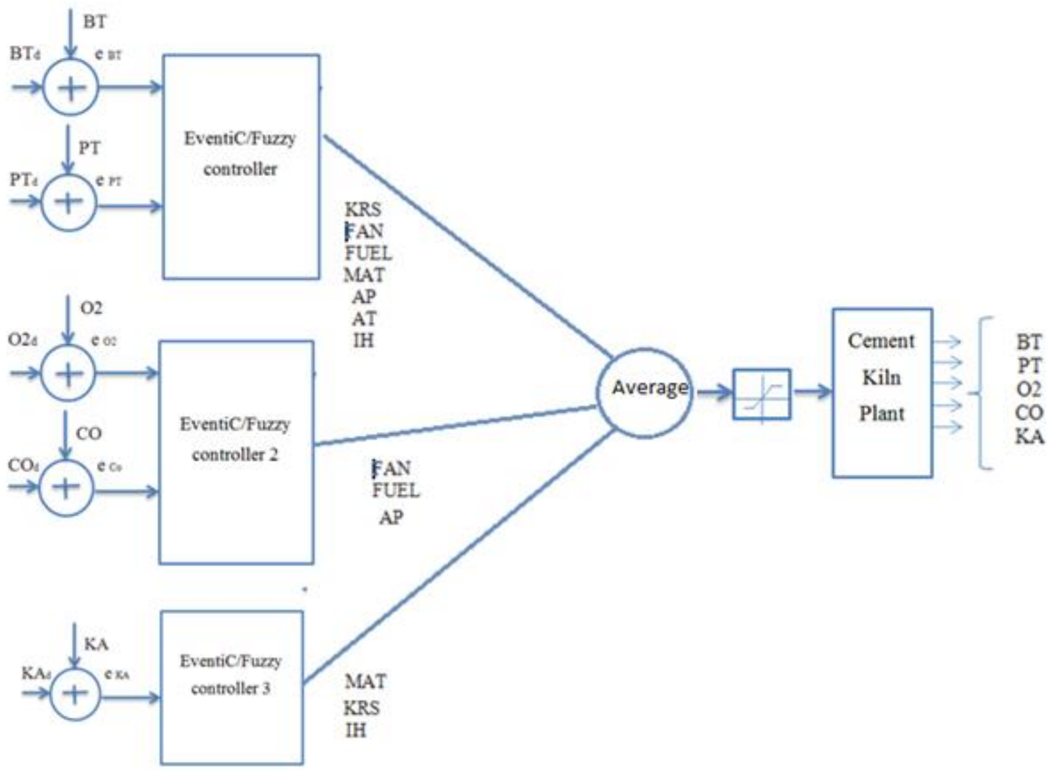


Figure 6.8. Segregated EventiC/fuzzy controllers integrated structure

Table 6.13. Authorized ranges for O_2, CO and KA variables

	O_2	**CO**	**KA**
Minimum	6	0	60
Maximum	25	0.6	170

Back-end and pre-heater temperature ranges are selected by experts in cement production and in our experiment have been presented in table 6.14.

Table 6.14. Selected ranges for BT and PT variables

	Back-end Temperature(°C)	**Pre-heater temperature(°C)**
Minimum	300	300
Maximum	700	600

6.4.4. Controller scheme design

6.4.4.1. Defining the controller's layers

The fuzzy controller consists of three layers which are:

Layer 1: Input layer

The input unit in this layer is the transformed process output error (e). This layer receives the error signal and uses a membership function to determine the relative contribution of the observed signals.

Layer 2: Rule Layer

The rule layer implements the link relating pre-conditions to consequences. Each rule has only one antecedent link from the input data.

Layer 3: Output Layer

All consequences are fully connected to outputs and are interpreted directly as the strength (weight) of the outputs. This layer performs centroid defuzzification to obtain inference output.

The presented controller is equivalent to a simplified adaptive fuzzy inference system (Mota et al., 1993) which integrates with the EventiC model, where layer 1 corresponds to the antecedent part of the fuzzy control rules, and layers 2 and 3 correspond to the conclusion part.

6.4.4.2. Input variable selection for cement quality and initialization point

EventiC algorithm sits between the cement plant's SCADA and fuzzy controller. It starts to be trained by reading the input and output events and calculations of SA weights over one month sampling per every minute. It also measures mean and standard deviations of inputs and outputs to build their membership functions. These membership functions are used for rule generation in IF–THEN form.

6.4.4.3. Defining the membership function

Defining the membership function is the second step of designing a fuzzy controller. Membership functions are defined based on the changes in set-points and errors.

A. Controlled Variables memberships

1. O_2 and CO contents and KA:

The measured values for O_2, CO and KA variables are compared with their authorized values (the range of table 6.13 is for good-quality cement). If the values are not in the normal range, the corresponding controller acts and generates the required control signals. The membership functions of CO, O_2 and KA variables are shown in figures. 6.9, 6.10 and 6.11.In the subspaces that the membership functions for these variables in their normal range are not defined, they will not have any role and interference for control. However, if one of these inputs goes out from its normal range, the corresponding controller overrides the control procedure.

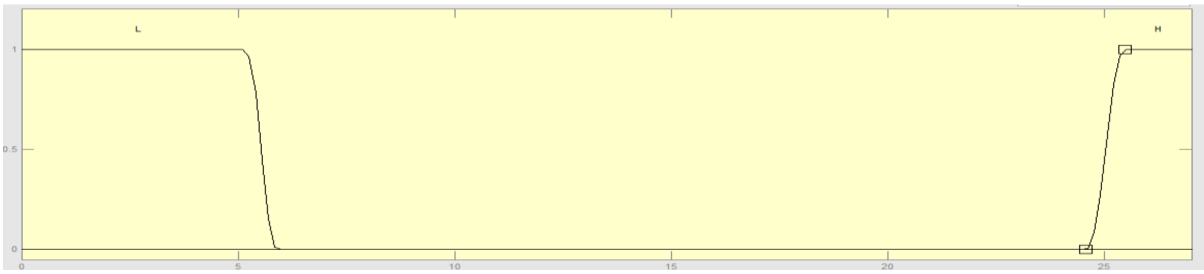


Figure 6.9. The membership functions for error of O_2variables

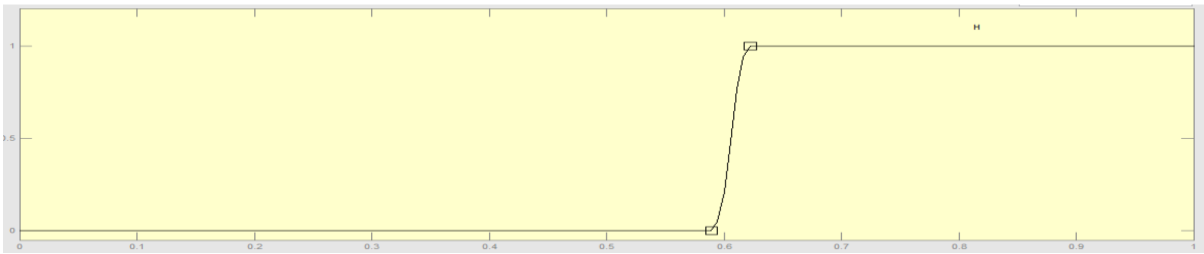


Figure 6.10.The membership functions for error of CO variables

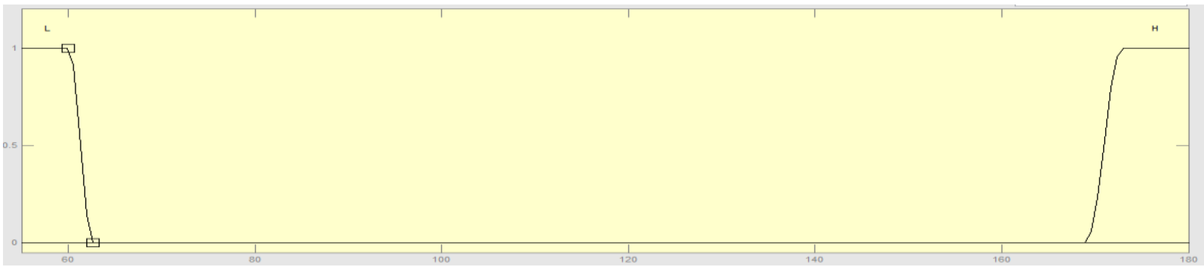


Figure 6.11.The membership functions for error of KA variables

2. BT and PT:

Three Gaussian membership functions have been considered for handling the error of BT and PT input variables. To find input variables' reference point, after the system meets its steady state, EventiC finds BT and PT variances and their means as reference points. These values are shown in table 6.15 in our experiment. Figure6.12shows the error of BT input variable membership function.

Table 6.15. BT and PT mean and variance after 1 month sampling

	Back-end Temperature	Pre-heater temperature
Reference point(Mean)	491	423
Variance	90	28

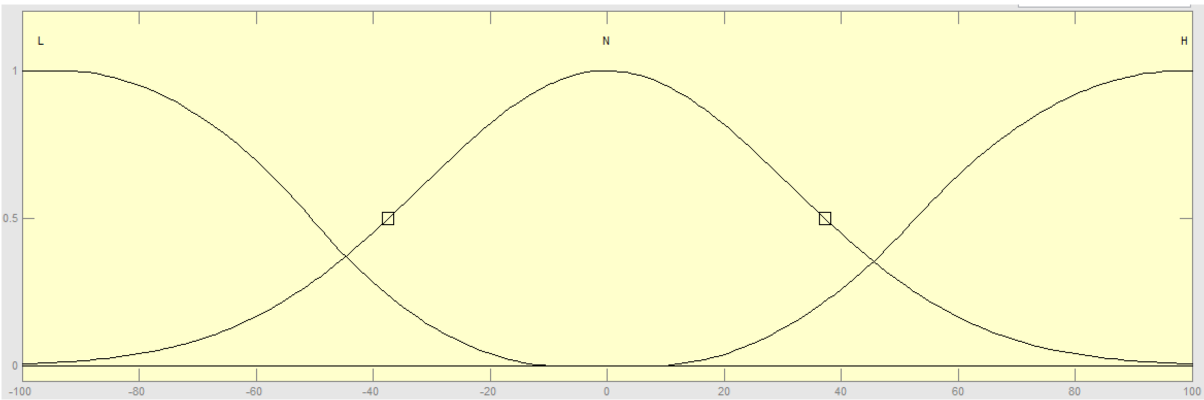


Figure 6.12. The membership functions for error of BT controlled variable

B. Control variables membership

Control variables membership functions as the kiln's input variables are measured by the EventiC algorithm and at a system steady state are measured and presented in table 6.16.

Table 6.16. Kiln control variables mean and variance after 1 month sampling

	KRS	FAN	FUEL	MAT	AP	IH	AT
Reference point(Mean)	2.7	30	5.3	14	37	6.51	1335
Variance	0.6	5	1	4	1	4	70

Figure 6.13 shows AP input variable membership function as an example between all inputs.

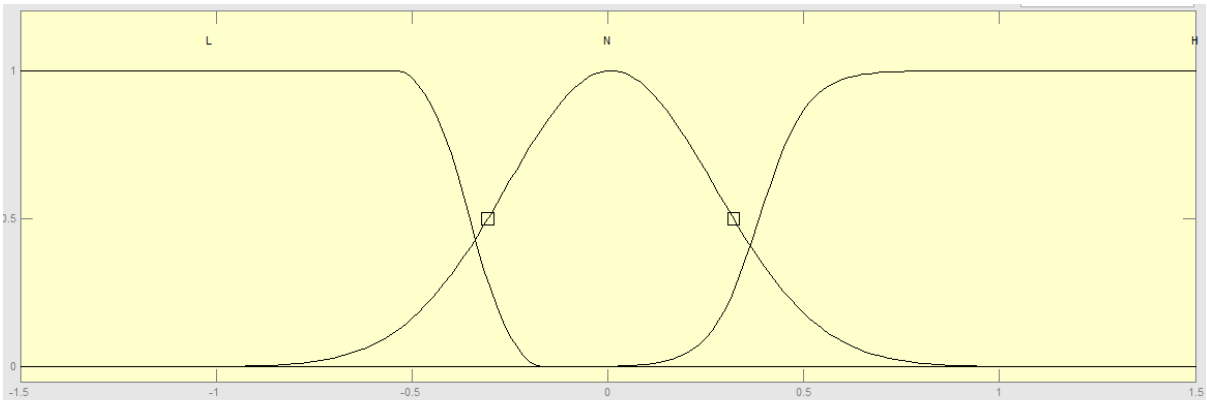


Figure 6.13. The membership functions for error of APcontrol variable

6.4.5.4. Control inference rules

Rules are generated by traversing the trained (at least 250 samples in the system's steady state) input/output weights as follows:

1. List all inputs that are known and have contributed to the outputs
2. Arrange the list by decreasing the absolute value of the weights.
3. Generate clauses for an IF–THEN rule from this ordered list.

Fuzzy controllers require substantial knowledge and experience from knowledge-based controllers (EventiC here) in order to keep the kiln operation smooth. Kiln control is normally the most intricate and challenging part of a cement plant's controls. Plants have to have the capability to immediately correct any errors that the plant expert controllers detect in the system. The control rules determined for the controller with 3 linguistic variables and 7 inputs are 3^7 =2187 rules.

Tuning fuzzy control rules by training data in our experiments required the setting of an initializing point. A group of training data is a pair of input-output data, in which the output data are desired output values, and the input data are relevant fuzzy input values. Such tuning data represents the skilled-operator control behaviour. It is proposed a tuning method for obtaining high performance FLCs by mean of EventiC whose components have been described.

6.5. Summary and conclusion

In this chapter, an adaptive FLC that modifies the fuzzy sets by using EventiC has been proposed. The objective is to determine the membership functions that produce maximum FLC performance according to the inference system. This method relies on having a set of weighted data against which the controller is tuned. Fuzzy rule extracted from EventiC algorithm. Fuzzy logic and EventiC possess contrasting application requirements. For example, fuzzy systems are appropriate if sufficient expert knowledge about the process is available, whilst the EventiC method is useful if sufficient process data is available or measurable. Both approaches build nonlinear systems based on bounded continuous variables, the difference being that EventiC systems are treated in a numeric quantitative manner, whereas fuzzy systems are treated in an symbolic qualitative manner.

Therefore, the integration of EventiC and fuzzy systems leads to a symbiotic relationship in which fuzzy systems provide a powerful framework for expert knowledge and representation, whilst EventiC provides learning capabilities and exceptional suitability for computationally efficient hardware implementations. The significance of this integration becomes even more apparent by considering their disparities. EventiC does not provide a strong scheme for knowledge representation, whilst fuzzy logic controllers do not possess capabilities for automated learning. An EventiC/fuzzy should be able to learn linguistic rules and/or membership functions, or optimize existing ones. Then, systems can start without rules, and create new rules until meeting the outputs.

In this chapter the advantages of EventiC/fuzzy techniques have been used to develop an intelligent control system for the cement kiln process. Cement production is a complex process, composed of a series of activities, and many variables which need to be manipulated and controlled. The EventiC/fuzzy controller is able to automatically extract all fuzzy parameters and design the structure of a FLC in order to control a nonlinear cement production process. The learning of the FLC is performed by the proposed EventiC, using the weight sets of input/output data, previously extracted by human experts or evolutionary algorithms. EventiC would reside on top of a typical data acquisition system (e.g. SCADA) and translate the data into cause-effect event models. It would link the set of events (EventiC inputs) to the set of performance variables (EventiC outputs). The fuzzy inference rules and parameters will be made available to the fuzzy modeller. Figure 6.14 illustrates a schematic of the proposed method in cement quality control.

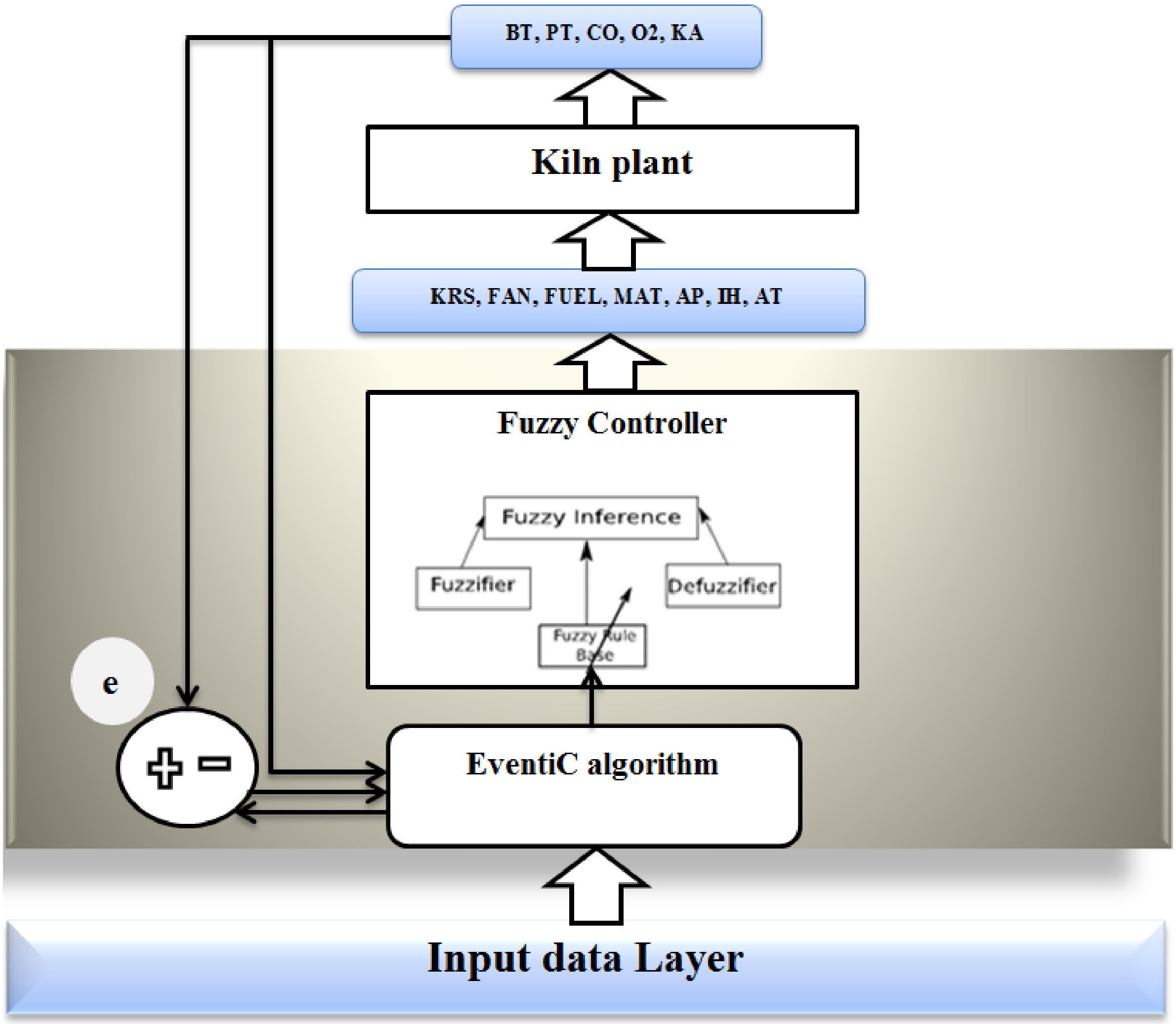


Figure 6.14. The EventiC/fuzzy controller structure

The proposed method does not require any prior knowledge concerning the fuzzy rule structure, membership function, implication and aggregation operators, defuzzification methods, or selection of adequate input variables. The main purpose of the EventiC application in FLCs is to develop a controller which constitutes a starting point for further adjustments. Additionally, the method may also be used to understand a process for which we have little or no information, as since it automatically extracts all fuzzy parameters; it is able to gather a knowledge-base about the process control. In order to validate the proposed methodology, it was applied to the cement production process.

7. Assessment of the Efficiency and Validity of the EventiC

The aim of this chapter is to validate and assess the efficiency of the proposed real-time EventiC sensitivity analysis techniques with EventTracker sensitivity analysis (Tavakoli,Mousavi & Broomhead, 2013b; Tavakoli, 2010). The objective of both techniques is to create an accurate representation of a system's state (the relationship between excitation and state) using real-time data. Both techniques are designed so that they can assemble sufficient knowledge about the internal/external eco-system of the causal relationships of events that will lead to a better understanding of systems' behaviour integrated with existing systems modelling techniques (e.g. stability or optimisation).The reason for choosing EventTracker over other SA methods is that the two are similar and comparable in being applicable to real-time data. Moreover, the two methods do not rely on the availability of statistically reliable or the homoscedasticity of historical data.

Chapter five's cement plant kiln experiment has been chosen to compare the similarity and dissimilarity of EventiC and EventTracker in suitability and applicability to industrial applications. At the beginning of comparison, the EventTracker sensitivity analysis technique will be reviewed concisely and then the results of both techniques will be analysed and compared.

7.1. EventTracker sensitivity analysis technique

EventTracker (Tavakoli, 2010; Tavakoli, Mousavi & Broomhead, 2013b)is a non-empirical causal relationship sensitivity analysis platform that relates field data to performance and process parameters. Raw data obtained from large data acquisition devices is monitored and its effect on system parameters is measured and reported to the higher layers of the information system.

In this way the EventTracker's method is able to construct a discrete event framework (Tavakoli, Mousavi & Komashie, 2008) where events are loosely coupled with respect to their triggers for the purpose of sensitivity analysis. The method has a clear advantage over the analytical and computational IVS method since it tries to understand and interpret the system state change in the shortest possible time with minimum computational overhead. The shortcomings of other SA methods were in their reliance on historical data and the generation of data samples which the system analyst then fits to known probabilistic equations. One method that was less reliant on analytical methods for extracting sensitivity indices was the entropy method. In this respect the entropy method was shown to be the closest technique but its shortcomings with regards to actual sensitivity, accuracy and computational efficiency became apparent when compared to the event tracking model, i.e. the 'EventTracker' in a real-world case study. The key principle of EventTracker is that every performance parameter can be affected by the total number of inputs to the system, by a degree. In effect, whatever happens in the environment and during the production process has a casual effect on the other, and on the total system. Thus, every measurable parameter within the system can be defined within the global set of inputs and the degree that affects that parameter. EventTracker in this sense can be considered to be a truly global sensitivity analysis method. The approach does not require any prior estimation of the data distribution (see fig.7.1).

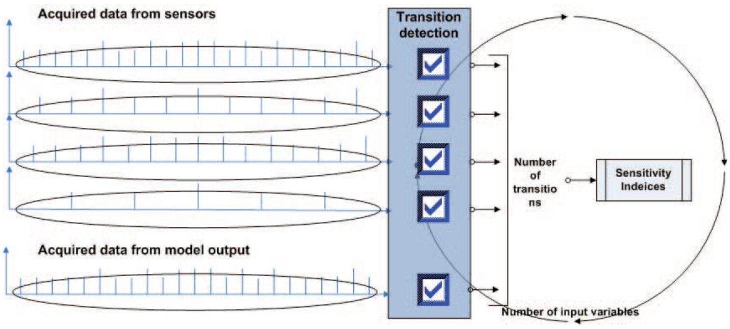


Figure 7.1. General view of the EventTracker method for sensitivity analysis

Tavakoli (Tavakoli, Mousavi & Broomhead, 2013b)designed various experiments such as a baking process and a refrigerator manufacturing process to compare the proposed event tracking sensitivity analysis method with a comparable method (that of entropy). An improvement of 10% in computational efficiency without loss in accuracy was observed. The comparison also showed that the time taken to perform the sensitivity analysis was 0.5% of that required when using the comparable entropy-based method.

The EventTracker technique has also been deployed in (Tavakoli, Mousavi & Poslad, 2013)a sensor-based data system in well-drilling system to solve time-critical dimensionality reduction problems with limited computational resources to control and optimize the deep drilling process (i.e. a complex component) to avoid disastrous malfunction of the complex drilling system.

7.1. The EventTracker algorithm

The EventTracker sequence algorithm follows a number of steps, which in brief, comprises the algorithm setting the parameters for Event Data and associated Event Trigger rules. The next step is to produce the input/output coincidence matrix based on the system specified scan rate. The third step is to extract the sensitivity indices of the parameters at specified intervals (set by system operators/engineers).The next step is to generate a normalised sensitivity index (SI) for each parameter in the analysis span (contiguous scan intervals). The penultimate step is to filter out the unimportant inputs by defining a cut-off threshold. For example, any input with less than 0.6 SI value is not influential enough on the specified output parameter. The final step is the validation and verification of results through a false-negative testing process.

7.2. A case study for comparing EventTracker and EventiC techniques

A major difference between EventTracker and EventiC techniques is fundamentally difference between their functionality. EventiC algorithm concept is based on clustering and grouping similar inputs and outputs in every time scale whilst EventTacker is analysing just an individual output correlation with inputs. EventiC designed to overcome the shortcoming of EventTracker (Tavakoli et al. 2010). Whilst EventTracker deals with 1 to many correlations, the EventiC cluster is intended to deal with many to many relationships. This main difference leads to a new clustering and grouping technique in system's input/output correlation area. As discussed in chapter 5, EventiC not only finds unknown/known relationships, but also it could find unknown/unknown relationship in system performance. It has the capability to identify group multiple causes to multiple effects.

To validate the EventiC sensitivity analysis technique, chapter five’s experiment has also been conducted with EventTracker sensitivity analysis. The reason for choosing EventTracker over other SA methods is the similarity of these two methods in working with real-time data and that neither have a reliance on the availability of statistically reliable or the homoscedasticity of historical data.

The results from deploying both methods over the existing 196 sensors and actuators in the SCADA of the cement manufacturing process are in table 7.1. The SA weights show that the sensitivity analysis results of both methods are very similar.

Table 7.1. Comparison of EventiC and Event Tracker's SA weight of a selected number of a kiln's input data over kiln production rate

Input Name	Sensitivity Level of kiln Production Rate with EventiC	Sensitivity Level of kiln Production Rate with EventTracker
Kiln temperature	**92 %**	**90%**
CO output	63 %	61%
I/h return in kiln	**90%**	**87%**
Kiln fan	**98%**	**95%**
CO_2 output	**97%**	**92%**
Motors pulls material from kiln	**92%**	**85%**
Injected O_2 to Kiln	37%	35%
Injected NO_2 to Kiln	54%	44%
Injected SO_2 to kiln	36%	36%

As has been discussed in section 5.2, the cut-off threshold filters out less important inputs with respect to their weights over the kiln output. This threshold is defined in both EventiC and EventTracker. An acceptable cut-off threshold that eliminates whole and less important inputs needs to conduct a false negative test. Figure 7.2 shows the minor differences in the cut-off (CT) thresholds which leads to the same filtered sensors. For instance, it shows that in EventiC with 90% CT, 18% of TDs are less important while in EventTracker with 87% CT, it finds 18% less important TDs. Experimentation revealed that for this industrial experiment on the EventiC algorithm, with a CT of 90%, with 18% of TDs (36 TDs) filtered out, the percentage of false negatives drops to 0.

The results also show that with the full 196 TDs, the EventiC algorithm took 40 seconds to calculate system output, whereas when using 160 TDs only 28 seconds were taken to achieve the same results i.e. a reduction of 30% in computation time. The computational time save for the EventTracker algorithm is 35%.

Measuring the computational effort on a personal computer with Intel®Core i7 CPU & 4.00 GB memory RAM, the average CPU utilization remains at 55% during the analysis run-time. The CPU utilisation using EventiC is10% better than EventTracker. Under the same experimental conditions, the CPU utilisation is 65%.

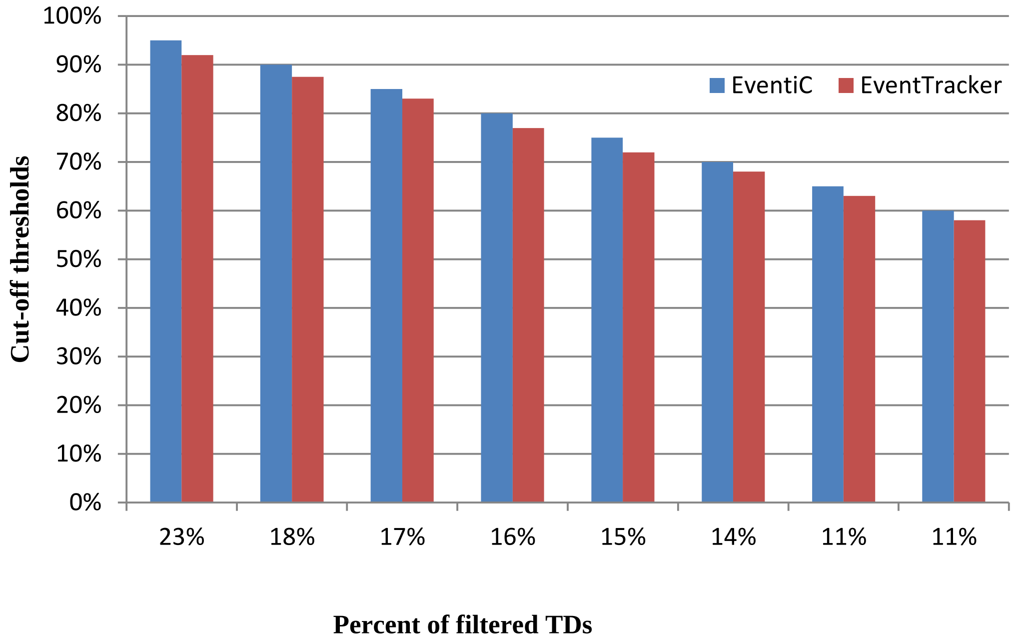


Figure 7.2. EventiC and EventTracker CT versus percent of filtered TDs

Table 7.2 summarises the similarities and differences of these two techniques in ten different categories including their major grouping/individual SA feature. The algorithm was run in MATLAB 2010b (7.11.0).

Table 7.2. Comparison table between EventiC and EventTracker techniques

Comparison Table	EventiC	Event Tracker
Grouping SA	**Yes**	**No**
Real-time sensitivity analysis	Yes	Yes
Applicable to large scale DAQ	Yes	Yes
Responsiveness to system volatilities	Yes	Yes
Complex and non-linear systems	Yes	Yes
Reliance on historical data/Statistical or model based equations	No	No
Computational complexities	No	No
CPU capacity	55%	65%
Algorithm speed	40 seconds	45 seconds
Computational time save	30%	35%

7.3. Conclusion

In this chapter a comparison experiment between EventiC and EventTracker was conducted to validate the proposed EventiC sensitivity analysis method. The rational for choosing EventTracker over other SA techniques is the similarity that exists between these techniques. EventTracker, like EventiC also does not rely upon the availability of statistically reliable data or the homoscedasticity of historical data.

These real-time SA techniques have been proposed to make systems more intelligent in dealing with real-time events and provide a more accurate representation of the system to the higher-level of mathematical formalism leading to intelligent controllers and decision making. These accurate real-time data engineers will increase precision and reduce the response time. Furthermore, these techniques remove all the logical boundaries of isolation that exist in complex systems with the principle that every acquirable knowledge or data (input) affects the output unless proven otherwise.

The logic behind the EventTracker and EventiC methods is the capture of the cause-effect relationship between input variables (triggers) and output variables (events) over given instances. The comparison was conducted on kiln processing in the cement industry. 196 of the existing sensors in the kiln were fed into the EventiC and EventTracker networks to minimize input dimensions and find the state of the system in real-time with respect to kiln productivity (kiln output). Results show that 18% of input sensors had little effect on kiln productivity in both techniques with a small difference in their CT results. However, the experiment confirmed that CPU capacity usage in EventTracker is approximately 10% more than the EventiC technique although the algorithm running speed was fairly equal.

8. Conclusion, contribution, and future work

This thesis proposes a technological capability which enabled us to acquire, exchange, and process data from any given physical or virtual system boundlessly. It fundamentally shifts our perceptions of systems boundaries. Our interlinked super systems need to respond to externally generated stimuli within a finite-specified period (i.e. Real-Time). With expansion and fluctuation of modern systems, producing time-critical accurate knowledge about the state of the system remains and continues to remain a major challenge for engineers.

The Hypothesis of this research was that: *"all the available knowledge about the internal and external events surrounding a defined system has an effect on its state, unless proven otherwise."* Within the context of the thesis, the key questions that this thesis intended to address were:

1. What would the implication of unbiased increase of input data and their potential relationships with one another and the system outputs be on our understanding of systems behaviour? Addressing the scalability problem.
2. How would the new knowledge about the new interrelationships between systems components allow us to better define performance (e.g. cost, quality, reliability, and fidelity)?
3. How would tracking and relating the events that represent the observable behaviour of the system lead to an increased insight to system functionalities and does it lead to more desirable improvements in the control of the system, and its optimal operation?

The first question posed in the research question was answered in Chapters 2, 3 and 4, where the definitions of system and methods of knowledge engineering and management was presented. Moreover, by designing the experiments through the systems simulation and observations on the live activities of a cement plant, the scalability of the approach was facilitated. Using real-time feed data from the plant and cross correlating the system inputs and outputs it was possible to assess the issues of significant increase in quantity and quality of data to monitor the performance of the system.

Building a powerful eco-system of causal events is one of the most natural approaches to de-cluttering complexity. Owing to the modern communication systems (i.e. Internet of Things), the engineering of Data and Knowledge of systems at near real-time speed is made possible. Our endeavour is to be able to understand (i.e. be able to explain), relate excitation with behaviour, to predict, to stabilize (Control), and to optimize (Operational Research) complex system at the shortest possible time.

In order to turn the theory into practice the author presents the concept of Event clustering (EventiC) as a platform for managing the interrelationships and internal dynamics of the components within the eco-system of embedded systems and their environment. It is designed to manage the causal relationships between the system and its operational environment as the system changes state and boundaries. EventiC can be used as a robust input variable selection technique for real-time systems. By equipping the embedded monitoring and control systems with the proposed technique a real-time data modelling platform emerges. This will enable control engineers and system designers to build more adaptable and responsive systems in pursuit of optimal performance and attain the stability of systems. This unique system will not only achieve significant improvements in the design of systems in its currently tested environment (i.e. process manufacturing), but has the potential to be applied to aerospace, automotive and smart metering applications. This new platform will be a foundation for sharing and integrating multiple users in various applications. With this capability it will be able to create cyber-physical systems that understand the impact of known and previously 'unknown' internal and external destabilizing factors and find quick solutions to stay functional and sustainable.

EventiC is not only an intelligent recorder of events, but an intelligent platform that enables preliminary data and knowledge construction. It is a complimentary middleware between plant and environmental information sources (raw data) with higher-level information management systems and optimization tools. Existing production control and management systems would benefit from the overlay of EventiC analysis within their existing data monitoring, reporting and analysis capabilities. Currently management in the production industry relies on known factors combined into predictive models, thus EventiC will be an instrument for detecting, classifying and analysing the impact of previously unknown factors.

The concept of real-time Event Clustering and data analysis, especially in the area of large scale data systems, is immensely important and the proposed method is ‘unique’. It is a foundation for dealing with the acquisition of large scale data and organising it in the form of linked interrelationships and clusters of relevance. This takes place at the lowest layer of interface between the physical system and the higher-level information framework. Thus, the proposed method should indeed be considered as the linking point between engineering and data modelling.

The key novelty of this research is its capability to acquire large sets of data and group the relevant input event data to the key performance indicators of a system and then rapidly generate an event-driven incidence matrix and measure the degree of influence of input sensors on outputs. Moreover, EventiC does not require prior knowledge of the analytical or statistical relationships that may exist between input and output variables. These accurate real-time data engineers will increase precision and reduce the response time. These techniques remove all the logical boundaries of isolation that exist in complex systems with the principle that every acquirable knowledge or data (input) affects the output unless proven otherwise.

The second question raised in this research question is addressed in chapters 5 and 6, where the state of the system or the output parameters of the system is defined by the Key Performance Indicators specified in the manufacturing industry. These indicators are product quality, productivity, production efficiency, resource utilisation, and inventory. They are represented by well-established transfer functions defined in manufacturing systems literature. The raw data emitting from the sensors and actuators in the plant integrated by a SCADA system represent the input parameters of the system.

The designed experiment enabled the simulation of acquisition of real-time data from the plant and conducting sensitivity analysis against the event that takes place during the production process. The observations took over 30 days at a rate of 1 minute sampling rate.

The sensitivities of all KPIs were assessed against all 196 available input data series. Whilst originally this was not the case. Individual KPIs were only connected to a pre-specified set of input series. Thus enabling us to assess the efficacy of de-modularity of the system. This resulted in improving the quality of input data, by reducing the number of data acquisition points by 18%. One needs to refer to the case study to recognise the challenges that a system with over 196 data acquisition points that generate large amount of data in short periods of time poses for the system controller. Reducing the amount of input data in close to real-time is a major achievement. Furthermore, the advantage of the proposed method is not only in filtering unwanted data, but in putting new sets of input data into play (another novel and unique feature) that was not considered relevant by the system control engineers at the outset of the design process, for example, in the operation of a kiln. It is believed that this will lead to superior mathematical formalism of problems and control systems, leading to better transfer functions or other analytical methods.

Finally to answer the third key question of the thesis, EventiC has been integrated into the cement kilns fuzzy controller to automate expert-knowledge based information in fuzzy controllers. The new modulations and extracted relationships between systems input and output parameters lead to better understanding of the system behaviour and more over reducing the lead time to return the system to optimal performance. This was proven by connecting resultant EventiC to the factory’s fuzzy controller, in which new membership functions showed to represent the dynamics of the system more effectively and accurately. EventiC/fuzzy controllers can obtain appropriate fuzzy controller inputs and create fuzzy control rule inference tables.

To validate the proposed EventiC sensitivity analysis method a comparison experiment between EventiC and EventTracker was conducted to ratify the results. The experiment confirmed that besides the event-clustering feature which is unique to EventiC, CPU capacity usage in EventTracker is approximately 10% more than with the EventiC technique but the algorithm running speeds were fairly equal.

8.1. Contribution to knowledge

This research has advanced the understanding of complex systems' internal dynamic and its interaction with environment, with particular reference to research key questions which leads to EventiC's technology in the following ways:

1. Propose a data and knowledge engineering technique to meet the challenges of the dynamic, autonomous, adaptive and self-organising embedded systems, and, seamless/secure interaction of the embedded system/cyber-physical systems with their environment.

2. Propose a technique to achieve a correlation analysis throughout the analysis span and deals with too many relationships of input and output parameters via grouping of relevant input-output event data by order of its importance in real-time.
4. Event clustering fundamentally shift our perceptions of systems boundaries with its quick ROC algorithm in processing the necessary information in near real-time which leads to better solutions for improving systems' performance (e.g. cost, quality, reliability, and fidelity).
3. Event clustering visualise system inputs and key performance (outputs) event incidence occurrences matrices and find appropriate corrective actions to maintain stability and optimal functionality externally generated stimuli.
4. Automatic extraction of the fuzzy controller parameters in integration with fuzzy controllers which leads to improvements in the control of the system, and its optimal operation.

The above contributions have been made by undertaking the following activities and realising the following achievements:

1. The introduction of Event-Clustering sensitivity analysis (EventiC) as a new paradigm in assembling the necessary knowledge (data analysis) for the purpose of systems modelling and control optimization. EventiC achieves this by simply (a) interpreting changes in the values of input-output (I/O) data as I/O events (b) detecting if the I/O events coincide, and (c) groups the I/O events, as related events. This (a)-(c) process happens at specified time intervals called scan rates. Scan rates can range from microseconds, to seconds, minutes, etc. Each scan registers a scenario of input-output, like a clip in a film. The weight of an input on output is calculated using the basic logic of the number of coincidences in a time span. Finally, and for the purpose of modelling, a return is made to the actual values of the I/O data. Such an approach could be considered a novel approach due to its process in understanding large-scale raw data. The automation of the preliminary data analysis has significantly reduced time-system modelling, design and validation.

2. With the implementation of the proposed technology into the SCADA system of a cement production plant of our industrial partner the EventiC technique was able to set

the production system to an optimal point, maximizing production rates whilst minimizing environmental impacts.

3. EventiC technology has also been integrated into the cement kilns fuzzy controller to automate some expert knowledge-based information in fuzzy controllers.

8.2. Future work

The opportunity to expand and build upon this thesis' findings exist in various aspects including, but not limited to, the following two approaches:

8.2.1. *Research which bounds to EventiC technology:*

- EventiC algorithm can be improved with much research on: definitions of instantaneous events, events with different delays and the scenarios of combination of events that lead to a specific output.
- Some steps have already been taken to customize EventiC applications for engine safety issues in the aerospace and the automotive industries to find key performance indicators. Furthermore, EventiC's application in the financial market could develop with regards to measuring oil industry key performance parameters with reference to price volatility.
- Introduce EventiC application for integrating the components of 'Internet of Things' (i.e. modern communication systems) at near real-time speed.
- The EventiC approach in sensitivity analysis can find the thresholds triggers for each input/output from the nature of the system.
- Improve and explore the application of the proposed technique in the demanding and complex environments of vehicular transport (aircraft, automobile and other vessels).

8.2.2. *Impact of the research*

- The new knowledge and insight vis-a'-vis the relationship between system input and output parameters may open new opportunities for formulating and reformulating dynamical systems behaviour,
- The production of industry based objective function models where key performance indicators can be combined to provide overall key performance factors of a system and then translate all in cost index.

8.3. Overall benefits of the research outcome

The outcome of the research is a novel approach for engineering data and greater understanding of the deterministic relationship between a system state and the dynamism of its internal and external functioning. This method is able to lay the foundation for robust real-time, predictive and optimized models. Through allowing a superior interpretation of the state of a system, EventiC is able to propose and establish a more efficient and fluent mechanism in the design and modelling of systems. This will be achieved by exploring the methods of establishing connectivity between (a) the network of embedded devices, (b) the network of the embedded system, and (c) the internet of things.

REFERENCES

Abu-Jamous, B., Fa, R., and Nandi, A. K.,2013. Biclustering. Integrative Cluster Analysis in Bioinformatics, pp. 265-282.

Ackoff, R.L., 2000. Creating the corporate future. Understanding Business Environments, Routledge, London, pp. 217-227.

Aggarwal, C. C., Zhai, C. (2012). A survey of text clustering algorithms. InMining Text Data (pp. 77-128). Springer US.

Aliev, R.A., Fazalollahi, B. and Vahidof, R.M., 2001. Genetic algorithm-based learning of fuzzy neural networks. Part 1: feed-forward fuzzy neural networks. Fuzzy Sets and Systems, 118(2), pp. 351-358.

Anderson, P.W., 1972. More is different. Science, New York, N.Y., 177(4047), pp. 393-396.

Andrieu, C., Doucet, A. and Holenstein, R., 2010. Particle markov chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 72(3), pp. 269-342.

Ankerst, M., Breunig, M. M., Kriegel, H. P., andSander, J., 1999. OPTICS: ordering points to identify the clustering structure. In *ACM Sigmod Record*, 28(2), pp. 49-60.

Armstrong, R.A., Slade, S. and Eperjesi, F., 2000. An introduction to analysis of variance (ANOVA) with special reference to data from clinical experiments in optometry. Ophthalmic and Physiological Optics, 20(3), pp. 235-241.

Arnold, M., Siemers, P., & Adamson, G., 2014. Cybernetics and contemporary philosophies of technology. In Norbert Wiener in the 21st Century (21CW), 2014 IEEE Conference on, pp. 1-7.

Ashby, W.R., 1964. Requisite variety. In An Introduction to Cybernetics (Chapter 11, pp. 202-218). London: Chapman & Hall Ltd.

Auder, B., Iooss, B., 2008. Global sensitivity analysis based on entropy. In Safety, reliability and risk analysis-Proceedings of the ESREL 2008 Conference.

Ayed, B. A., Halima, B. M., Alimi, A. M., 2014. Survey on clustering methods: Towards fuzzy clustering for big data. IEEE *Soft Computing and Pattern Recognition (SoCPaR), 6th International Conference of*, pp. 331-336.

Baldwin, C. Y., Clark, K. B., 2003. Managing in an age of modularity. *Managing in the Modular Age: Architectures, Networks, and Organizations*, pp.*149*.

Banks, D. L., Olszewski, R. T., & Maxion, R. A., 2003. Comparing methods for multivariate nonparametric regression. Communications in Statistics-Simulation and Computation, 32(2), pp. 541-571.

Barajas, S.E., Reyes, C.A., 2005. Your fuzzy relational neural network parameters optimization with a genetic algorithm, Fuzzy Systems, 2005. FUZZ'05. The 14th IEEE International Conference on 2005, IEEE, pp. 684-689.

Battelle, 2002. Toward a Sustainable Cement Industry, Substudy 8: Climate Change, Humphreys K., and Mahasenan, M., Battelle Memorial Institute.

Ben Ayed, A., Ben Halima, M., & Alimi, A. M., 2014. Survey on clustering methods: Towards fuzzy clustering for big data. 2014 6th International Conference of In Soft Computing and Pattern Recognition (SoCPaR), pp. 331-336.

Beylkin, G., Kurcz, C.,d Monzon, L., 2008. Fast algorithms for Helmholtz Green's functions. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science, 464(2100), pp. 3301-3326.

Bishop, M., Christopher M., 2006. Pattern Recognition and Machine Learning. Springer.
Blake, Merz C.J., UCI repository of machine learning databases.

Blanchard, B.S., 2004. System engineering management. John Wiley & Sons.
BLUM, A.L. and LANGLEY, P., 1997. Selection of relevant features and examples in machine learning. Artificial Intelligence, 97(1), pp. 245-271.

Blessing, L. T., Chakrabarti, A., 2009. *DRM: A Design Research Methodology*. Springer London.

Blum, A. L., Langley, P., 1997. Selection of relevant features and examples in machine learning. *Artificial intelligence*, *97*(1), pp. 245-271.

Bo, Y., Yi, L., & Shouning, Q., 1997, A rule-based cement kiln control system using neural networks. In Intelligent Processing Systems, 1997. ICIPS'97. 1997 IEEE International Conference on, Vol. 1, pp. 493-497.

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A., 2013. A review of feature selection methods on synthetic data. Knowledge and information systems, 34(3), pp.483-519.

Bolton, W., 2008. Mechatronics: electronic control systems in mechanical and electrical engineering. Pearson Education.

Borgonovo, E. and Peccalti, L., 2007. Global sensitivity analysis in inventory management. International Journal of Production Economics, 108(1), pp. 302-313.

Box GEP, Wilson KB. , 1951. On the experimental attainment of optimum conditions. J R Stat Soc;Series B(13),pp.1–45

Boulding, K., 1956.General Systems Theory, The Skeleton of Science Management Science,

Buonomo, A. and Lo Schiavo, A., 2010. Nonlinear distortion analysis via perturbation method. International Journal of Circuit Theory and Applications, 38(5), pp. 515-526.

Calderon, M. L., 2010. The Design Research Methodology as a Framework for the Development of a Tool for Engineering Design Education. In DS 62: Proceedings of E&PDE 2010, the 12th International Conference on Engineering and Product Design Education-When Design Education and Design Research meet..., Trondheim, Norway, 02.-03.09.

Carmona, C. J., González, P., del Jesus, M. J., & Herrera, F., 2010, NMEEF-SD: non-dominated multi objective evolutionary algorithm for extracting fuzzy rules in subgroup discovery", IEEE Transactions on Fuzzy Systems, 18(5), pp. 958-970.

Caruana, R., De sa, V.R., 2003. Benefitting from the variables that variable selection discards. The Journal of Machine Learning Research, 3, pp. 1245-1264.

Cassandras, C. G., Lafortune, S., 2008. Introduction to discrete event systems. Springer Science & Business Media.

Cembureau, 1999. Cement Production, Consumption Data: World Cement Market in Figures 1913–1995, World Statistical Reviews 19 and 20. CEMBUREAU—The European Cement Association, Brussels, Belgium.

Cement Sustainability Initiative, 2015. Word Business Council for Sustainable Development. http://www.wbcsdcement.org/index.php/en/about-cement/cement-production

Chan, A. P., 2004. Key performance indicators for measuring construction success. *Benchmarking: an international journal*, *11*(2), pp. 203-221.

Chen, Y., Wu, K., Chen, X., Tang, C., Zhu, Q., 2014. An entropy-based uncertainty measurement approach in neighborhood systems. Information Sciences, 279, pp. 239-250.

Cloke, H., Pappenberger, F. and Renaud, J., 2008. Multi-method global sensitivity analysis (MMGSA) for modelling floodplain hydrological processes. Hydrological Processes, 22(11), pp. 1660-1674.

Cohen, J., Cohen, P., West, S.G. and Aiken, L.S., 2013. Applied multiple regression/correlation analysis for the behavioral sciences. Routledge.

Conesa, J.A., Galvez, A., Malteos, F., Martin-Gullon, I. and Font, R., 2008. Organic and inorganic pollutants from cement kiln stack feeding alternative fuels. Journal of hazardous materials, 158(2), pp. 585-592.

Cukier, R.I., Fortuin, C.M., Shuler, K.E., Petschek, A.G., Schaibly, J.H., 1973. Study of sensitivity of coupled reaction systems to uncertainties in rate coefficients. 1. Theory, J. Chem. Phys., 59, (8), pp. 3873– 3878.

Cusumano, M.A., 1997, How Microsoft makes large teams work like small team. Sloan Management Review, pp. 9-20.

Caruana,R. 2003.Benefiting from the Variables that Variable Selection Discards ,Journal of Machine Learning Research 3 ,pp. 1245-1264.

Dai, J.,Wang, H., Zheng, J.,Ge, Q., 2014. Temperature sensitivity of plant phonology in temperate and subtropical regions of China from 1850 to 2009. *International Journal of Climatology*35:10.1002/joc.2015.35.issue-6, 913-922. Online publication date: 1-Feb-2015.

Damineli, B. L., Kemeid, F. M., Aguiar, P. S., & John, V. M., 2010. Measuring the eco-efficiency of cement use. Cement and Concrete Composites, 32(8), pp.555-562.

Danishvar,M., Mousavi, A., Sousa, P. and Araujo, R., 2013. Event-clustering for real-time data modelling, Automation Science and Engineering (CASE), 2013 IEEE International Conference on 2013, IEEE, pp. 362-367.

Danishvar,M., Mousavi,A.,Sousa,P., 2014. EventiC for improved real time Data Modelling and Sensitivity Analysis, IEEE Multi-conference in Systems and Control (MSC 2014) ,Anitibas ,France, October 2014 .

Danishvar,M., Mousavi,A.,Broomhead,P., 2014. The Real-Time Unaware Event Data Clustering (EventiC): the Eco-System of Causality in Complex Systems, Submmited for IEEE Industrial Informatic 2015

Delgado, M.R., Nagai, E.Y. and De arruda, Luciavaleriaramos, 2009. A neuro-coevolutionary genetic fuzzy system to design soft sensors. Soft computing, 13(5), pp. 481-495.

Devedzic, V., 1995, Knowledge-based control of rotary kiln. In Industrial Automation and Control: Emerging Technologies, 1995, International IEEE/IAS Conference on, pp. 452-458.

Distefano,S., Puliafito, A.,Kishor S. Trivedi, 2012.Dynamic aspects and behaviours of complex systems in performance and reliability assessment. SIGMETRICS Performance Evaluation Review, 39(4), pp.71-78.

Dimitriosevangelopulos, 2012. Process optimisation in the cement industry; the kiln problem, Brunel University.

Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci (PNAS),* Volume 95, pp. 14863-14868.

Elghazel, H., Aussem, A., 2013. Unsupervised feature selection with ensemble learning. *Machine Learning*, *98*(1-2), pp.157-180.

Engin, T. and Ari, V., 2005. Energy auditing and recovery for dry type cement rotary kiln systems—A case study. Energy Conversion and Management, 46(4), pp. 551-562.

Fahad, A., Alashtari, N., Tari, Z., Alamri, A.Khalil,I., Zomaya,A. and Bouras, A. ,2014, A Survey of Clustering Algorithms for Big Data: Taxonomy & Empirical Analysis. IEEE Transactions on emerging topics in computing, 2(3).pp.267-279.

Fallahpour, M., Fatehi, A., Araabi, B.N. and Azizi, M., 2007. A supervisory fuzzy control of back-end temperature of rotary cement kilns. Control, Automation.

Fallahpour, M., Fatehi, A., Araabi, B.N. and Azizi, M.., 2008. A neurofuzzy Controller for Rotory cement Kiln, the 17th World Congress the International Federation of Automatic ControlSeoul, Korea.

Farahmand, H., Rashidnejad, M., Mousavi, A., Gharaveisi, A., Irving, M. and Taylor, G., 2012. Hybrid mutation particle swarm optimisation method for available transfer capability enhancement. International Journal of Electrical Power & Energy Systems, 42(1), pp. 240-249.

Feng, S., Banerjee, R., & Choi, Y., 2012. Syntactic stylometry for deception detection. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Association for Computational Linguistics, 2, pp. 171-175.

Fidaros, D., Baxevanou, C., Dritselis, C., Vlachos, N., 2007. Numerical modelling of flow and transport processes in a calciner for cement production. Powder Technology, 171(2), pp. 81-95.

Fock, E., 2014. Global sensitivity analysis approach for input selection and system identification purposes-a new framework for feed forward neural networks. IEEE transactions on neural networks and learning systems, 25(8) , pp. 1484-1495.

Forrester, J.W, 1958. Industrial Dynamics: A Major Breakthrough for Decision Makers", *Harvard Business Review*, 36(4), pp. 37-66.

Garud, R., Jain, S., Kumaraswamy, A., 1999. An integrative perspective for managing knowledge. *Cognition*, *Knowledge and Organizations*, *6*, pp.195-214.

Ghanem, R.,Ghiocel, D., 1996. A Comparative Analysis of FORM/SORM and Polynomial Chaos Expansions for Highly Nonlinear Systems, Engineering Mechanics (1996) 1996, ASCE, pp. 535-538.

Gheyas, I.A., Smith, L.S., 2010. Feature subset selection in large dimensionality domains. Pattern Recognit. 43, pp.5–13.

Ghosh, S. N., 2003. *Advances in cement technology: chemistry, manufacture and testing*. CRC Press.

Ghosh, T., Dan, P.K., 2011, effective clustering method for group technology problems: A short communication, e-Journal of Science & Technology, 6(4).

Glower, J. S., Munighan, J., 1997. Designing fuzzy controllers from a variable structures standpoint. *Fuzzy Systems, IEEE Transactions on*, *5*(1), pp.138-144.

Granger, C. W., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society, pp. 424-438.

Griensven, A.V, Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., & Srinivasan, R., 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models. Journal of hydrology, 324(1), pp.10-23.

Groover, M.P., 2007. Automation, production systems, and computer-integrated manufacturing. Prentice Hall Press.

Groover,M.P., 2014, Automation, Production Systems, and Computer-Integrated Manufacturing ;4th Ed, Pearson Education: Singapore.

Haken, H., 1977. *Synergetic. An Introduction. Nonequilibrium Phase Transitions and Self-organization in Physics, Chemistry, and Biology*. Berlin.

Hand, D.J., Mannila, H. and smyth, P., 2001. Principles of data mining. MIT press. London.

Helbing, D., 2007. Managing complexity: insights, concepts, applications. Springer Publishing Company, Incorporated.

Hassan, Y. F., Habeb, N., 2012. Hybrid system of PCA, rough sets and neural networks for dimensionality reduction and classification in human face recognition. IJIIP: International Journal of Intelligent Information Processing, 3(1), pp.16-24.

Helbig , D. ,2008.Managing Complexity: Insights, Concepts, Applications. Springer.

Helton, J.C., 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. Reliability Engineering & System Safety, 42(2), pp. 327-367.

Hempel, S., Koseska, A., Kurths, J. and Nikoloski, Z., 2011. Inner composition alignment for inferring directed networks from short time series. Physical Review Letters, 107(5), pp. 054101.

Herrera, F., Lozano, M. and VERDEGAY, J.L., 1995. Tuning fuzzy logic controllers by genetic algorithms. International Journal of Approximate Reasoning, 12(3), pp. 299-315.

Hoff, A. J., 2012. Advanced EPR: applications in biology and biochemistry. Elsevier.

Hoff, J., 2013. The Analogical Turn: Rethinking Modernity with Nicholas of Cusa Part II: Nocholas of Cusa and the genealogy of modern concepts of Space and Autonomy", William B Eerdmans Publishing Company, Grand Rapids, Michigan / Cambridge UK, 2012.

Holland, J. H., 1975. Adaptation in natural and artificial systems, Ann arbor, MI: university of Michigan press.

Hooker, C.A., 1981. Towards a general theory of reduction. Part I: Historical and scientific setting. Dialogue, 20(01), pp. 38-59.

Hoppner, F., Klawonn, F., 2000. Obtaining interpretable fuzzy models from fuzzy clustering and fuzzy regression, Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000. Proceedings. Fourth International Conference on 2000, IEEE, pp. 162-165.

Hora, S.C., Helton, J.C., 2003. A distribution-free test for the relationship between model input and output when using Latin hypercube sampling. Reliability Engineering & System Safety, 79(3), pp. 333-339.

Hruschka, E. R., Campello, R. J. G. B., Freitas, A. A., De Carvalho, A. P. L. F. , 2009. A survey of evolutionary algorithms for clustering. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 39(2), pp.133-155.

Hu, Z., Shi, P., 2010. Sensitivity analysis for biomedical models, IEEE Trans. on Medical Imaging, 29(11) , pp. 1870-1881.

Hung, J.Y., Gao, W., HUNG, J.C., 1993. Variable structure control: a survey. Industrial Electronics, IEEE Transactions on, 40(1), pp. 2-22.

INCOSE. 2015. The Challenge of Complex Systems. [ONLINE] Available at: http://www.incose-coa.org/the-challenge-of-complex-syst. [Accessed 01 October 2015].

INCOSE/ Sheard,S. , Cook, S., 2015. A Complexity Primer for Systems Engineers. [ONLINE] Available at: http://www.aiaa.org/uploadedFiles/Events/Conferences/2013_Conferences/2013_-_Aviation/Detailed_Program/Complexity_Primer_draft.pdf. [Accessed 01 October 2015].

INCOSE, 2015, Systems Engineering Handbook: A Guide for System Life Cycle Processes and Activities', version 4.0. Hoboken, NJ, USA: John Wiley and Sons, Inc, ISBN: 978-1-118-99940-0

Ingalls, B., 2008. Sensitivity analysis: from model parameters to system behaviour. Essays Biochem, 45, pp.177-194.

ISO/IEC. 2009. "Systems and Software Engineering Vocabulary (SEVocab)" - ISO/IEC 24765. in International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) [database online]. Geneva, Switzerland, 2009 [cited September 20 2015]. Available fromhttp://pascal.computer.org/sev_display/index.action.

Isukapalli, S.S., 1999. Uncertainty analysis of transport-transformation models.

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31(8), pp. 651-666.

Jamshidi, M., 1997. Fuzzy control of complex systems. Soft Computing, 1(1), pp.42-56.

Jin, Y., Yue, H., Brown, M., Liang, Y., Kell, D.B., 2007. Improving data fitting of a signal transduction model by global sensitivity analysis, American Control Conference, 2007. ACC'07 2007, IEEE, pp. 2708-2713.

Juang, C. and Lu, C., 2005. Fuzzy Controller Design by Hybrid Evolutionary Learning Algorithms, Fuzzy Systems, 2005. FUZZ'05. The 14th IEEE International Conference on 2005, IEEE, pp. 525-529.

Kabir, G., Abubakar, A. and El-Nafaty, U., 2010. Energy audit and conservation opportunities for pre-processing unit of a typical dry process cement plant. Energy, 35(3), pp. 1237-1243.

Karypis, G., Han, E. H., & Kumar, V., 1999. Chameleon: Hierarchical clustering using dynamic modelling. *Computer, 32*(8), 68-75.

Keating, C., Rogers, R., Unal, R., Dryer, D., Sousa-Poza, A., Safford, R., Peterson, W. and Rabdi, G., 2003. System of Systems Engineering. Engineering Management Journal, 15(3).

King, J.R., 1980a. Machine-component grouping in production flow analysis: an approach using a rank order clustering algorithm. International Journal of Production Research, 18(2), pp. 213-232.

King, J.R., 1980b. Machine-component group formation in group technology. Omega, 8(2), pp. 193-199.

Klir, G.J., 1970. An approach to general systems theory. Van Nortrand, Princeton, NJ,.

Kohavi, R. and John, G.H., 1997. Wrappers for feature subset selection. Artificial Intelligence, 97(1), pp. 273-324.

Kohavi, R., John, G. H., 1997. Wrappers for feature subset selection. Artificial intelligence, 97(1), pp.273-324.

Krieger, T.J., Durston, C., and Albright, D.C., 1977. Statistical Determination of Effective Variables in Sensitivity Analysis', Trans. Am. Nuc. Soc. 28, pp.515-516.

Krishnapuram, R. and Ferg, C., 1992. Fitting an unknown number of lines and planes to image data through compatible cluster merging. Pattern Recognition, 25(4), pp. 385-400.

Krzykacz-Hausmann, B., 2006. An approximate sensitivity analysis of results from Complex computer models in the presence of epistemic and aleatory uncertainties. Reliability Engineering & System Safety, 91(10), pp. 1210-1218.

Krzykacz-Hausmann, B., 2001. Epistemic sensitivity analysis based on the concept of entropy, International symposium on sensitivity analysis of model output 2001, pp. 53-57.

Kung, S. Y., 2015, Visualization of big data. In *Cognitive Informatics & Cognitive Computing (ICCI* CC), 2015 IEEE 14th International Conference on*, pp. 447-448.

Langlois, R. N., 2002. Modularity in technology and organization. Journal of economic behaviour & organization, 49(1), pp.19-37.

Laplante, P.A., 1993. REAL-TIIME SYSTEMS DESIGN AND ANALYSIS.

Laszlo, A., Krippner, S., 1998. Systems theories: Their origins, foundations, and development. ADVANCES IN PSYCHOLOGY-AMSTERDAM-, 126, pp. 47-76.

Laszlo, E., 1972. The relevance of general systems theory.

Ledeneva, Y., 2006. Automatic estimation of parameters to reduce rule base of fuzzy control complex systems. a master thesis in august.

Lewin,R., 1993. Complexity: Life at the Edge of Chaos. Maxwell Macmillan International.

Liao, W. K., Liu, Y., and Choudhary, A., 2004. A grid-based clustering algorithm using adaptive mesh refinement. In 7th Workshop on Mining Scientific and Engineering Datasets of SIAM International Conference on Data Mining, pp. 61-69.

Liu, Y., Yousuff hussaini, M. and Ökten, G., 2013. Optimization of a Monte Carlo variance reduction method based on sensitivity derivatives. Applied Numerical Mathematics, 72, pp. 160-171.

Livingston, S. A., Antal, J. , 2009, A case of inconsistent equating: How the man with four watches decides what time it is. Applied Measurement in Education, 23(1), pp.49-62.

Lorenz, E., 2000. The Butterfly Effect. The chaos avant-garde: Memories of the early days of chaos theory, 39, p.91.

Lu, Y., Cohen, I., Zhou, X. S., & Tian, Q. ,2007. Feature selection using principal feature analysis. In Proceedings of the 15th international conference on Multimedia , pp. 301-304.

MacQueen,J.,1967, Some methods for classification and analysis of multivariate observations, in Proc. 5th Berkeley Symp. Math. Statist. Probab., Berkeley, CA, USA, pp. 281–297.

Madhanagopal, R., Chandrasekaran, R., 2014, Selecting Appropriate Variables for DEA Using Genetic Algorithm (GA) Search Procedure. International Journal of Data Envelopment Analysis and Operations Research, 1(2) , pp. 28-33.

Mahasenan, N., Smith, S., Humphreys, K. and Kaya, Y., 2003. The cement industry and global climate change: current and potential future cement industry CO_2 emissions, Greenhouse Gas Control Technologies-6th International Conference. Oxford: Pergamon 2003, pp. 995-1000.

Malhi, A., & Gao, R. X. , 2004. PCA-based feature selection scheme for machine defect classification. Instrumentation and Measurement, IEEE Transactions on, 53(6), pp.1517-1525.

Makaremi, I., Faltehi, A. Araabi, B.N., 2008. Lipschitz numbers: a medium for delay estimation, 17th IFAC world congress, Seoul, Korea 2008.

Mamdani, E.H. and Assilian, S., 1975. An experiment in linguistic synthesis with a fuzzy logic controller. International journal of man-machine studies, 7(1), pp. 1-13.

Marino, S., Hogue, I.B., Ray, C.J., Kirschner, D.E. , 2008. A methodology for performing global uncertainty and sensitivity analysis in systems biology, J.Theor. Biol., 254(1), pp. 178–196.

Maruyama, M., 1963. On the Force Equivalents of Dynamical Elastic Dislocations with Reference to the Earthquake Mechanism, Journal of Earthquake research institute ,41, pp.467-486 .

May, R., Dandy, G., Maier, H., 2011, Review of input variable selection methods for artificial neural networks. Artificial neural networks—methodological advances and biomedical applications , pp. 19-44.

May, R. M. , 1976. Simple mathematical models with very complicated dynamics. Nature, 261(5560), pp.459-467.

McKay, M. D., Beckman, R. J., Conover, W. J., 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21(2), pp.239-245.

Mcrae, G.J., Tilden, J.W., Seinfeld, J.H., 1982. Global sensitivity analysis—a computational implementation of the Fourier amplitude sensitivity test (FAST). Computers & Chemical Engineering, 6(1), pp. 15-25.

Mendes, J., Seco, R., Araujo, R., 2011. Automatic extraction of the fuzzy control system for industrial processes, Emerging Technologies & Factory Automation (ETFA), 2011 IEEE 16th Conference on 2011, IEEE, pp. 1-8.

Miguel, M.S., Johnson, J.H., Kertesz, J., Kaksi, K., Diaz-Guilera, A., Machkay, R.S., LORETO, V., ÉRDI, P., HELBING, D., 2012. Challenges in complex systems science. arXiv preprint arXiv:1204.4928.

Mikkola, J. H. , 2003. Modularity, component outsourcing, and inter-firm learning. *r&d Management*, *33*(4), pp.439-454.

Miller, J. G., 1978. Living Systems. New York: McGraw-Hill.pp.1102.

Minnebo, W., Stijven, S. and Vladislaveva, K.,2007. Empowering Knowledge Computing with Variable Selection..Online publication date: 1-Sep-2014. http://styven.be/thesis/ThesisWouterSean_v2.pdf

Mitra, S. and Hayasi, Y., 2000. Neuro-fuzzy rule generation: survey in soft computing framework. Neural Networks, IEEE Transactions on, 11(3), pp. 748-768.

Morisawa, S., Inoue, Y., 1974. On the selection of a ground disposal site by sensitivity analysis. Health physics, 26(3), pp. 251-261.

Mota, A.M., Fonesca, J.A. and Sanots, D.M., 1993. An adaptive control system for an industrial tunnel kiln, Industrial Electronics, 1993. Conference Proceedings, ISIE'93-Budapest., IEEE International Symposium on 1993, IEEE, pp. 102-105.

Mousavi, A., Hamdi, M., Sarhadi, M., 2006. Quick-response decision-making in the food processing industry. *International Journal of Industrial and Systems Engineering*, *2*(1), pp.99-122.

Naranjo, M., Brownlow, D.T. and Garza, A., 2011. CO_2 capture and sequestration in the cement industry. Energy Procedia, 4, pp. 2716-2723.

Ndou, V., Schina, L., Passiante, G., Del Vecchio, P., De Maggio, M., 2010. Toward an Open Network Business Approach. In: Proceedings of the 4th IEEE International Conference on Digital Ecosystems and Technologies, Dubai, 13–16 April 2010, pp. 282–287.

Nelles, O., 2001. Nonlinear system identification: from classical approaches to neural networks and fuzzy models. Springer.

Noshirvani, G., Fatehi, A.,Arrabi, B., Shirvani, M. Azizi, M., 2009. Comparison of rotary cement kiln identified models, Control and Automation, 2009. ICCA 2009. IEEE International Conference on 2009, IEEE, pp. 1290-1295.

Page, S. E., 2009. Understanding Complexity. The Great Courses. Chantilly, VA, USA: The Teaching Company.

Pena, J. M., Lozano, J. A. & Larranaga, P., 1999. An empirical comparison of four initialization methods for the K-Means algorithm. *Pattern Recognition Letters*, 20(10), pp. 1027-1040.

Prigogine,I. , 1995. Science of chaos or chaos in sciences: a rearguards battle. Physicalia Magazine, 17, pp. 213-18.

Rabitz, H., Alis, ÖF., 1999. General foundations of high-dimensional model representations. Journal of Mathematical Chemistry, 25(2-3), pp. 197-233.

Ramadge, P.J., Wonham, W.M., 1989. The control of discrete event systems. Proceedings of the IEEE, 77(1), pp. 81-98.

Reznik, L.,1997. Fuzzy controller's handbook: how to design them, how they work. Newnes.

Reznik, L., Ghanayem, O. and Bourmistrov, A., 2000. PID plus fuzzy controller structures as a design base for industrial applications. Engineering Applications of Artificial Intelligence, 13(4), pp. 419-430.

Rojas, I., Pomares, H., Ortega, J. and Prieto, A., 2000. Self-organized fuzzy system generation from training examples. Fuzzy Systems, IEEE Transactions on, 8(1), pp. 23-36.

Rutkowski,L.,2007.Clustering for data mining: A data recovery approach.Psychometrika, 72(1), pp.109-110.

Ruby Jr, C. W., 1997, A new approach to expert kiln control. In Cement Industry Technical Conference. 1997. XXXIX Conference Record., 1997 IEEE/PCA, pp. 399-412.

Saidur, R., Hossain, M., Islam, M., Fayaz, H. and Mohammed, H., 2011. A review on kiln system modeling. Renewable and Sustainable Energy Reviews, 15(5), pp. 2487-2500.

Saltelli, A., Annoni, P., 2010. How to avoid a perfunctory sensitivity analysis, Environmental Modeling and Software 25, pp.1508-1517

Saltelli, A., Bolado, R., 1998. An alternative way to compute Fourier amplitude sensitivity test (FAST). Computational Statistics & Data Analysis, 26(4), pp.445-460.

Saltelli, A., 2002. Sensitivity analysis for importance assessment. Risk Analysis, 22(3), pp. 579-590.

Saltelli, A., Tarantola, S. and Chan, K., 1999. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics, 41(1), pp. 39-56.

Sanchez, R., & Mahoney, J. T., 1996. Modularity, flexibility, and knowledge management in product and organization design. *Strategic management journal*, *17*(S2), PP. 63-76.

Seising, R., 2010. Cybernetics, system (s) theory, information theory and fuzzy sets and systems in the 1950s and 1960s. Information Sciences, 180(23), pp. 4459-4476.

Shannon, T.R., 1989. An introduction to the world-system perspective. Westview Press Boulder, CO.

Shannon, C. E., 1948. A mathematical theory of communication, *Bell System Tech. Journal 27*, pp. 379--423.

Shi, Y., Mizumoto, M., 2001. An improvement of neuro-fuzzy learning algorithm for tuning fuzzy rules. Fuzzy Sets and Systems, 118(2), pp. 339-350.

Simon, H., 2003. The architecture of complexity. Managing in the modular age: architectures, networks, and organizations, , pp. 411.

Sillitto, H. G., 2009. "On Systems architects and systems architecting: some thoughts on explaining the art and science of system architecting" Proc of INCOSE IS 2009, Singapore, 20-23 July

Sobol, I.M, 1990. Sensitivity estimates for nonlinear mathematical models. Matematicheskoe Modelirovanie, 2 (1), pp.112-118.

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 55(1-3), pp. 271-280.

Sugeno, M., 1985. An introductory survey of fuzzy control. *Information sciences*, *36*(1), pp.59-83.

Tabakhi, S., Moradi, P., Akhlaghian, F., 2014. An unsupervised feature selection algorithm based on ant colony optimization. Engineering Applications of Artificial Intelligence, 32, pp.112-123.

Takagi, T. and Sugeno, M., 1985. Fuzzy identification of systems and its applications to modeling and control. Systems, Man and Cybernetics, IEEE Transactions on, (1), pp. 116-132.

Tavakoli, S., Mousavi, A., Poslad, S., 2013. Input variable selection in time-critical knowledge integration applications: A review, analysis, and recommendation paper. Advanced Engineering Informatics, 27(4), pp. 519-536.

Tavakoli, S., Mousavi, A. and Broomhead, P., 2013. Event tracking for real-time unaware sensitivity analysis (EventTracker). Knowledge and Data Engineering, IEEE Transactions on, 25(2), pp. 348-359.

Tavakoli, S., 2010. Ageneric predictive information system for resource planning and optimisation.

Tavakoli, S., Mousavi, A., Komashie, A., 2008. A generic framework for real-time discrete event simulation (DES) modelling, Proceedings of the 40th Conference on Winter Simulation 2008, Winter Simulation Conference, pp. 1931-1938.

Taylor, Harry FW. Cement chemistry. Thomas Telford, 1997.

Ünler, A., Gungor, Z., 2009. Applying K-harmonic means clustering to the part-machine classification problem. Expert Systems with Applications, 36(2), pp. 1179-1194.

Uysal, I., Güvenir, H. A. (1999). An overview of regression techniques for knowledge discovery. The Knowledge Engineering Review, 14(04), 319-340.

Vakharia, A.J., 1986. Methods of cell formation in group technology: A framework for evaluation. Journal of Operations Management, 6(3), pp. 257-271.

Van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., Srinivasan, R., 2006.A global sensitivity analysis tool for the parameters of multi-variable catchment models. Journal of hydrology, 324(1), pp. 10-23.

Vega-Pons, S. & Ruiz-Shulcloper, J., 2011. A survey of clustering ensemble algorithms. *Int J Pattern Recognit Artif Intell,* 25(3), pp. 337-372.

Volkova, E., Iooss, B., Van Dorpe, F., 2008. Global sensitivity analysis for a numerical model of radionuclide migration from the RRC “Kurchatov Institute” radwaste disposal site. Stochastic Environmental Research and Risk Assessment, 22(1), pp. 17-31.

Von bertalanfey, L., 1956. General system theory. General systems, 1(1), pp. 11-17.

Von bertalanfey, L., 1968. General Systems Theory: Foundations, Development, Applications. George Braziller.

Von bertalanfey, L., 1972. The history and status of general systems theory. Academy of Management Journal, 15(4), pp. 407-426.

Von Neumann, J., Burks, A. W. ,1966. Theory of self-reproducing automata. *IEEE Transactions on Neural Networks*, *5*(1), pp. 3-14.

Waldrop, M.,1993. *Complexity: The emerging science at the edge of order and chaos*. Simon and Schuster.

Wang, J. Z, Li, J., 2003. Automatic linguistic indexing of pictures by a statistical modeling approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *25*(9), pp.1075-1088.

Wang, L., 1994. Adaptive fuzzy systems and control: design and stability analysis. Prentice-Hall, Inc.

Wang, L., Mendel, J.M., 1992. Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. Neural Networks, IEEE Transactions on, 3(5), pp. 807-814.

Wang, J., Zhou,S., Yi,Y., Kong,J., 2014. An Improved Feature Selection Based on Effective Range for Classification. The Scientific World Journal 2014, pp.1-8.

Watanabe, K., Tang, J., Nakamura, M., Koga, S., Fukuda, T., 1996. A fuzzy-Gaussian neural network and its application to mobile robot control. Control Systems Technology, IEEE Transactions on, 4(2), pp. 193-199.

WBCSD, I. 2009. Cement Technology Roadmap: Carbon Emissions Reductions up to 2050. *World Business Council for Sustainable Development and International Energy Agency.< http://wbcsdcement. Org/pdf/technology/WBCSD-IEA_Cement% 20Roadmap. Pdf*.

Weaver, W., 1949. Recent contributions to the mathematical theory of communication. The mathematical theory of communication, 1.

Wiener, N., 1948 (A). Cybernetics. Hermann Paris.

Wiener, N., 1948 (B). Cybernetics, or Communication and Control in the Animal and the Machine. *New York: ffiley*.

Wiener, N.,1949, A New Concept of Communication Engineering, in: Electronics, pp. 74–76.

Wiener, N., Schadé, J. P., 1965. *The Cybernetics of the Nervous System*. Elsevier Publicarion.17.

Wiener, N., 1965. Cybernetics: or the Control and Communication in the Animal and the Machine, 25.

Xiao, X. et al., 2003. *Gene clustering using self-organizing maps and particle swarm optimization.* Indianapolis, s.n., pp. 154-163.

Yager, R., Filev, D. P., 1994. Generation of fuzzy rules by mountain clustering. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology*, *2*(3), pp.209-219.

Yao, Z., Li, X., Wang, Z., 2010. Research and application of neural network PID control in cement industry. In *Industrial and Information Systems (IIS), 2010 2nd International Conference on*, 2, pp. 424-427.

Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E., Ruzzo, W. L., 2001. Model-based clustering and data transformations for gene expression data. Bioinformatics, 17(10), pp.977-987.

Yu,L., Liu, H., 2004. Efficient feature selection via analysis of relevance and redundancy. J. Mach. Learn. Res., 5, pp. 1205–1224.

Yu, W., Wilson, D., Young, B., Harris, T., 2009. Variance decomposition of nonlinear systems. In Control and Automation, ICCA 2009. IEEE International Conference on, pp. 738-744.

Zadeh, L.A., 1973. Outline of a new approach to the analysis of complex systems and decision processes. Systems, Man and Cybernetics, IEEE Transactions on, (1), pp. 28-44.

Zadeh, L.A., 1965. Fuzzy sets. Information and control, 8(3), pp. 338-353.

Zadeh, L.A., 1952. Selection of mathematics and engineering some basic problems in communication of information. Transactions of the New York Academy of Sciences, 14(5 Series II), pp. 201-204.

Zaheeruddin, Z., Anwer, J., 2005. A Simple Technique for Generation and Minimization of Fuzzy Rules, Fuzzy Systems, 2005. FUZZ'05. The 14th IEEE International Conference on 2005, IEEE, pp. 489-494.

Zeeman, E. C., 1977. *Catastrophe theory: Selected papers*, Addison-Wesley, *pp.1972–1977*.

Zeigler, B., 1999. Support for hierarchical modular component-based model construction in DEVS/HLA. Proceedings of 1999 Spring Simulation Interoperability Workshop, 99S-SIW-066.

Zeigler, B.P., Praehofer, H. and Kim, T.G., 2000. Theory of modelling and simulation: integrating discrete event and continuous complex dynamic systems. Academic press.

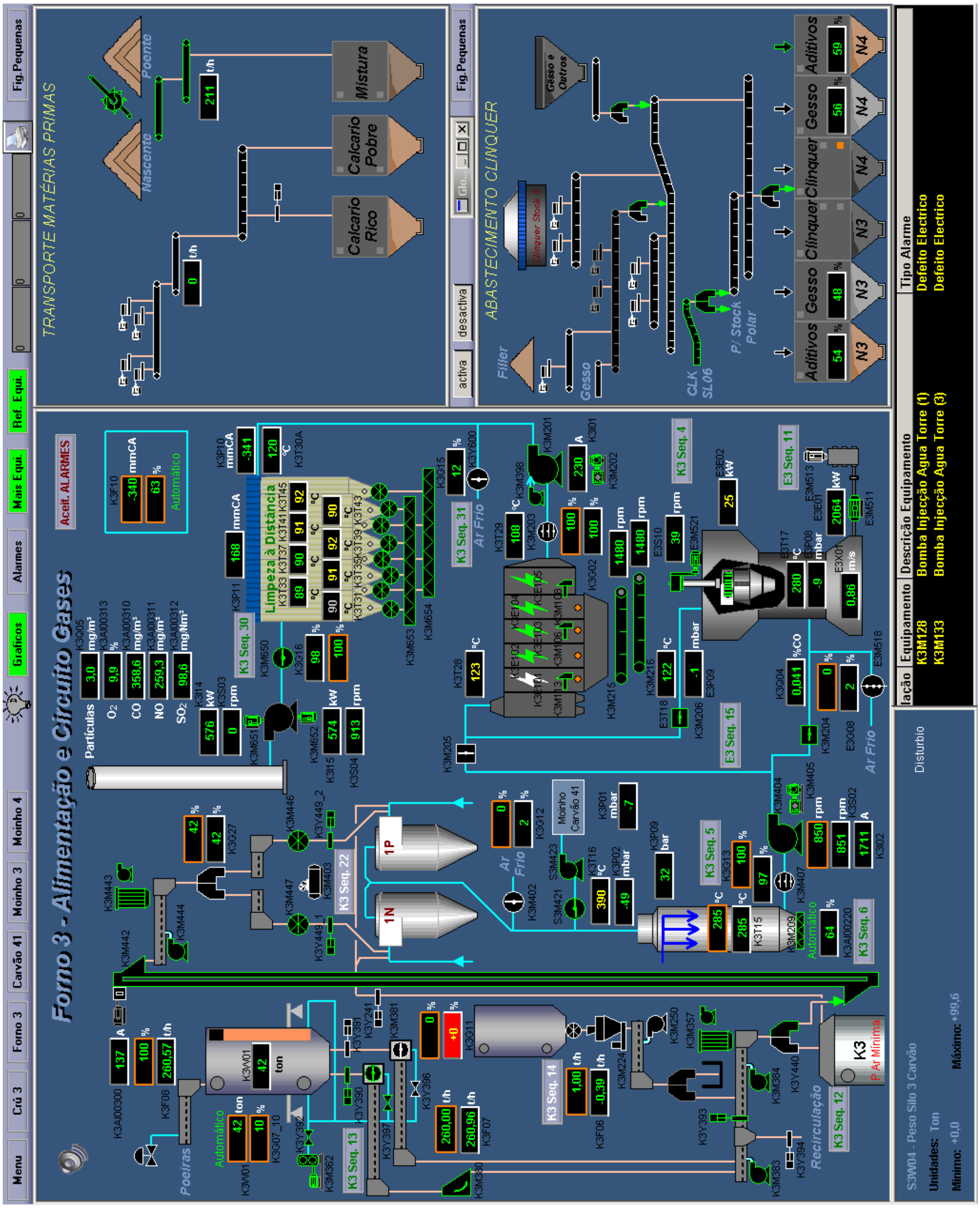
Zheng, Y., Rundell, A., 2006. Comparative study of parameter sensitivity analyses of the TCR-activated Erk-MAPK signalling pathway, IEE Proc. Syst. Biol., 153, (4), pp. 201 –211

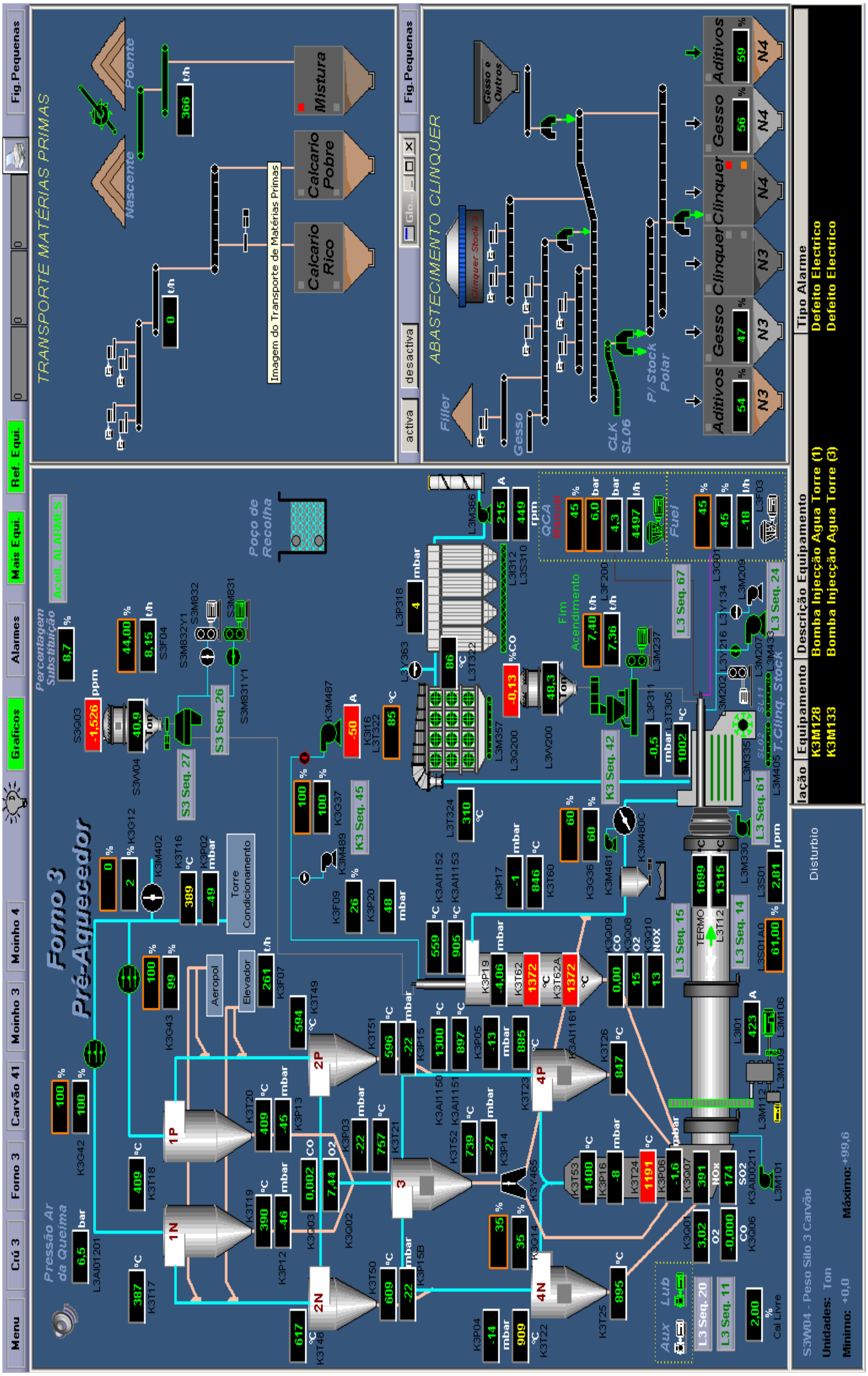
Zhu, Y., 2001. Multivariable system identification for process control. Elsevier.

Zi, Z., 2011. Sensitivity analysis approaches applied to systems biology models. Systems Biology, IET, 5(6), pp. 336-346.

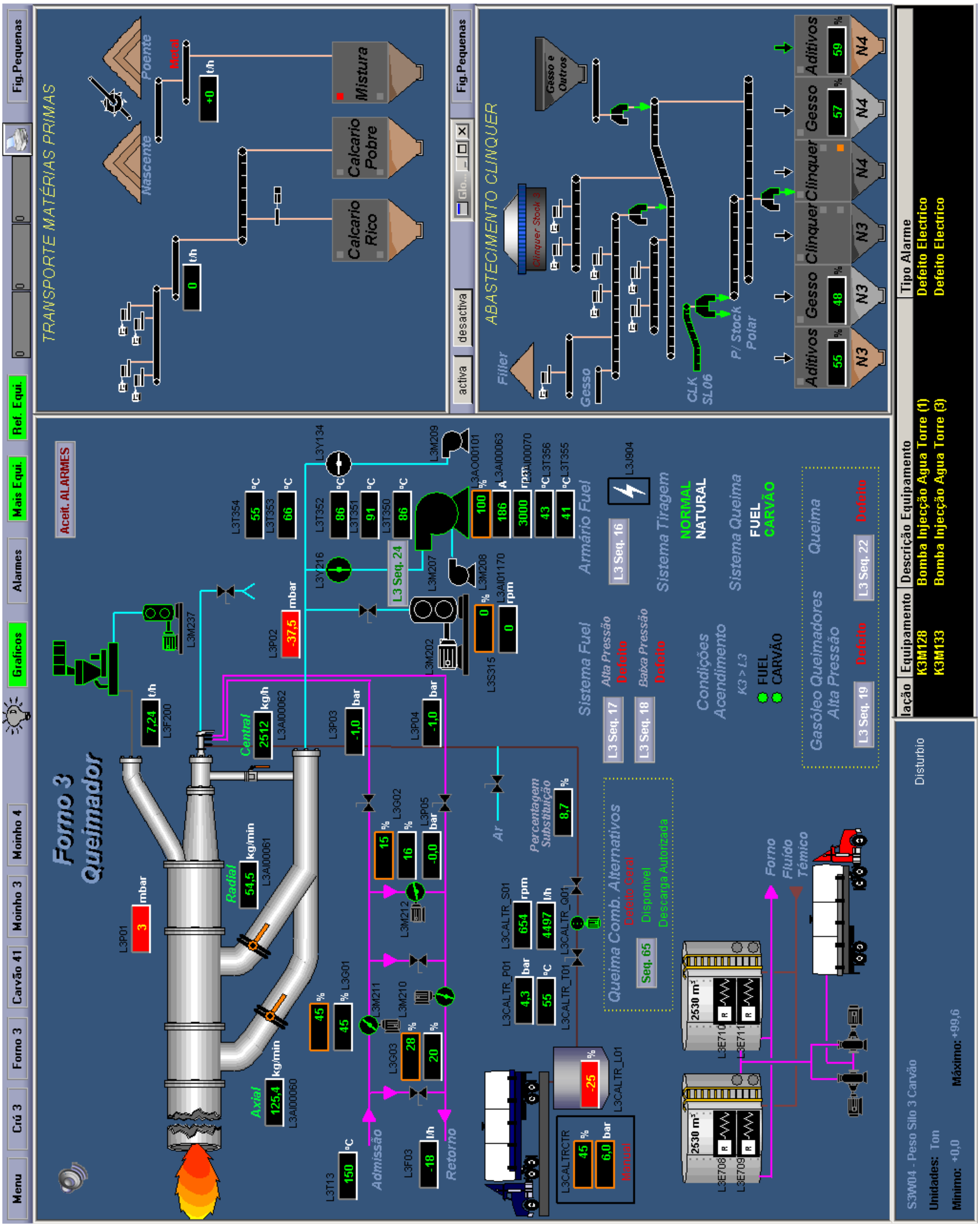
Appendices

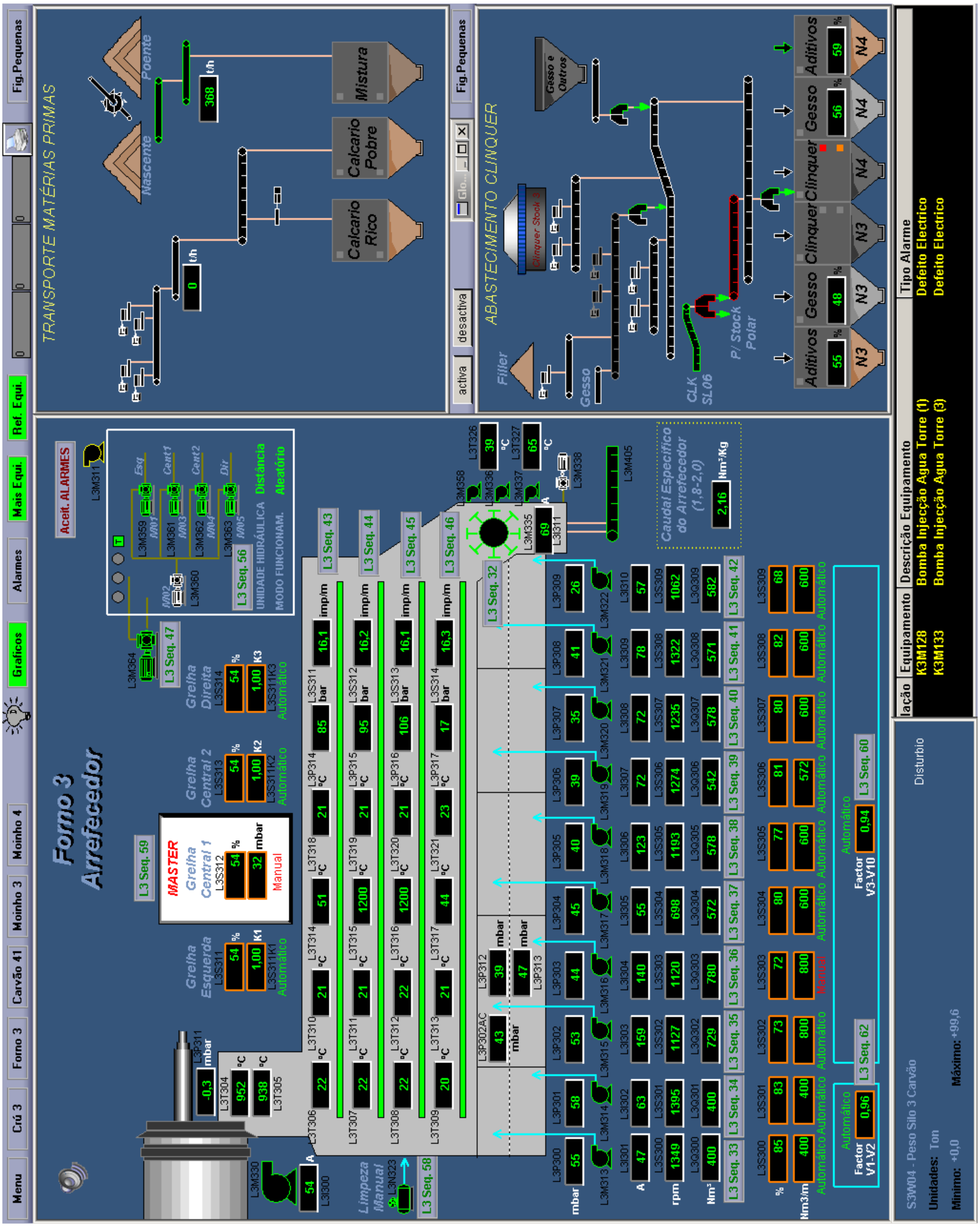
Appendix A: Schematic of our industrial partner's SCADA output including the list of sensor names

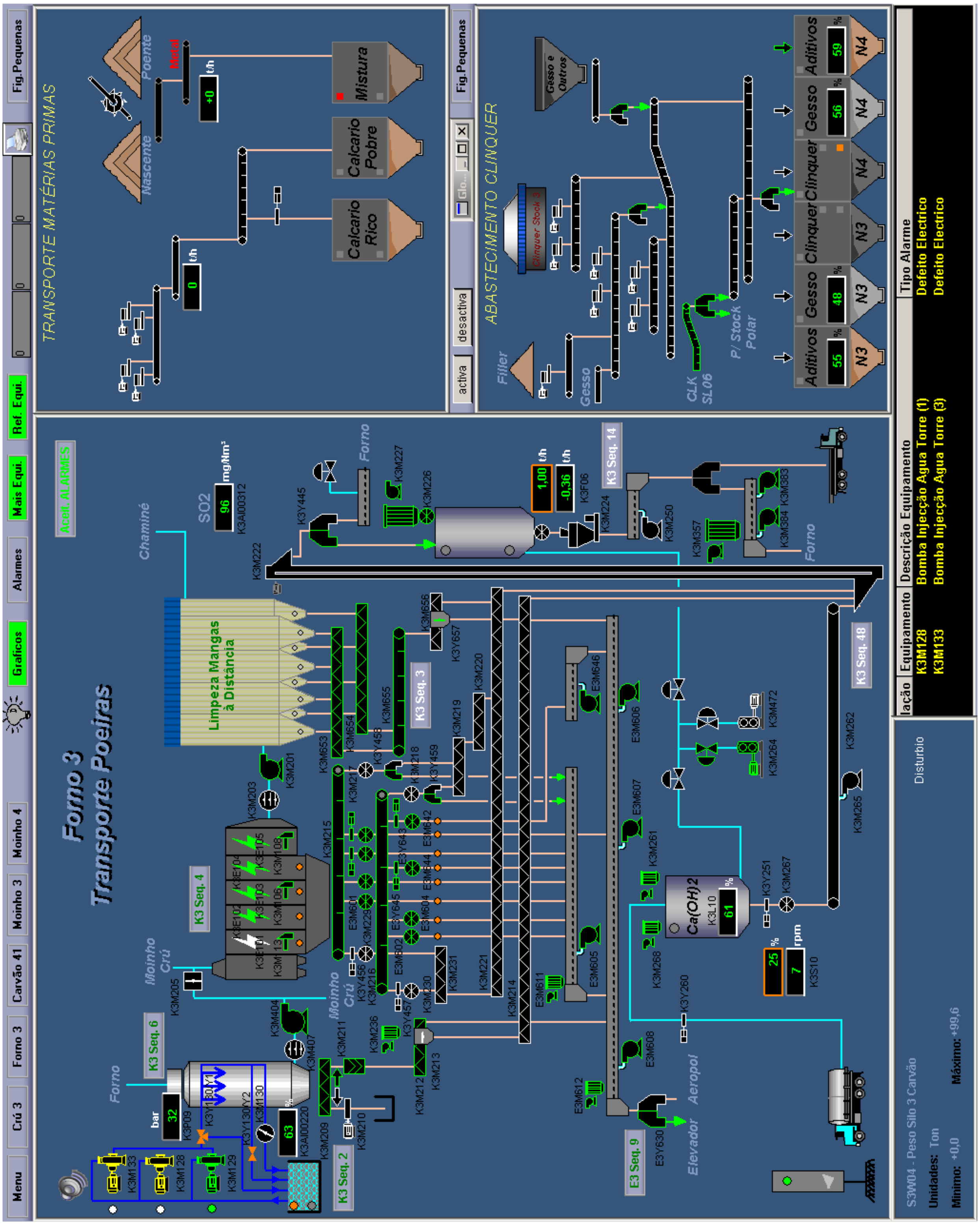


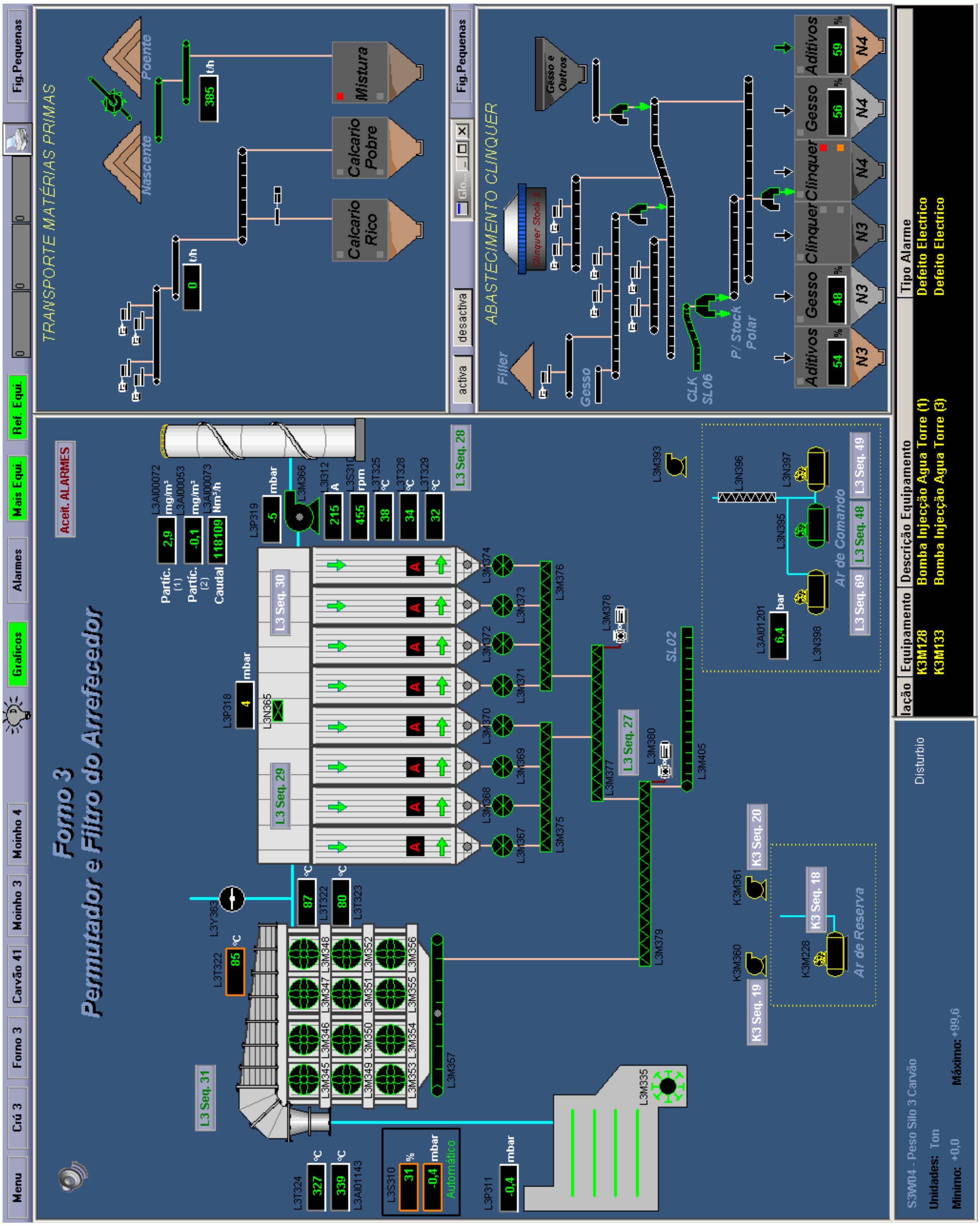


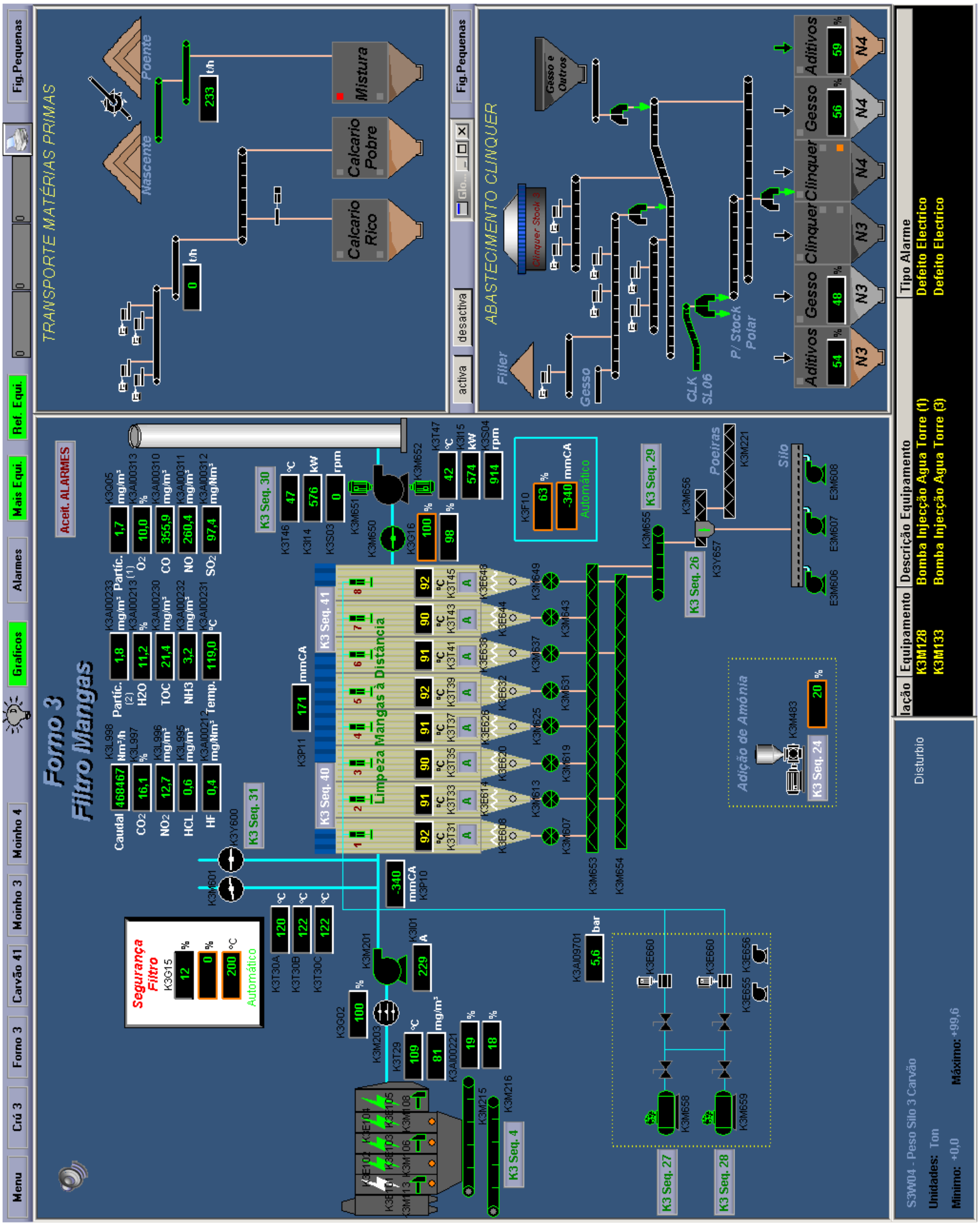
Menu
Crú 3
Forno 3
Carvão 41
Moinho 3
Moinho 4
Alarmes
Gráficos
Mais Equi.
Ref. Equi.
Fig.Pequenas
Forno 3
Pré-Aquecedor
Pressão Ar da Queima
Percentagem Substituição
Acert. ALARMES
Poço de Recolha
Torre Condicionamento
Aeropol
Elevador
TRANSPORTE MATÉRIAS PRIMAS
Nascente
Poente
Mistura
Calcario Pobre
Calcario Rico
Imagem do Transporte de Matérias Primas
ABASTECIMENTO CLINQUER
activa
desactiva
Filler
Gesso
CLK SL06
P/ Stock Polar
Gesso e Outros
Aditivos
Gesso
Clinquer
Clinquer
Gesso
Aditivos
N3
N3
N3
N4
N4
N4
Fuel
Disturbio
S3W04 - Peso Silo 3 Carvão
Unidades: Ton
Mínimo: +0,0
Máximo: +99,6
Equipamento
Descrição Equipamento
Tipo Alarme
K3M128
Bomba Injecção Agua Torre (1)
Defeito Electrico
K3M133
Bomba Injecção Agua Torre (3)
Defeito Electrico


Forno 3
Queimador
TRANSPORTE MATÉRIAS PRIMAS
ABASTECIMENTO CLINQUER


Menu
Crú 3
Forno 3
Carvão 41
Moinho 3
Moinho 4
Graficos
Alarmes
Mais Equi.
Ref. Equi.
Fig.Pequenas
Forno 3
Arrefecedor
Aceit. ALARMES
TRANSPORTE MATÉRIAS PRIMAS
Calcario Rico
Calcario Pobre
Mistura
Nascente
Poente
ABASTECIMENTO CLINQUER
Filler
Gesso
Gesso e Outros
Clinquer Stock
P/ Stock Polar
CLK SL06
Aditivos
Gesso
Clinquer
N3
N4
Greiha Esquerda
MASTER
Greiha Central 1
Greiha Central 2
Greiha Direita
Limpeza Manual
Caudal Especifico do Arrefecedor (1.8-2.0)
Equipamento
Descrição Equipamento
Tipo Alarme
K3M128
Bomba Injecção Agua Torre (1)
Defeito Electrico
K3M133
Bomba Injecção Agua Torre (3)
Defeito Electrico
S3W04 - Peso Silo 3 Carvão
Disturbio


Menu
Crú 3
Forno 3
Carvão 41
Moinho 3
Moinho 4
Graficos
Alarmes
Mais Equi.
Ref. Equi.
Fig.Pequenas
Forno 3
Transporte Poeiras
Limpeza Mangas à Distância
Chaminé
SO2
Acnt. ALARMES
TRANSPORTE MATÉRIAS PRIMAS
Calcario Rico
Calcario Pobre
Mistura
Nascente
Poente
ABASTECIMENTO CLINQUER
activa
desactiva
Equipamento
Descrição Equipamento
Tipo Alarme
K3M128
Bomba Injecção Agua Torre (1)
Defeito Electrico
K3M133
Bomba Injecção Agua Torre (3)
Defeito Electrico
S3W04 - Peso Silo 3 Carvão
Unidades: Ton
Minimo: +0,0
Máximo: +99,6
Disturbio


Menu
Crú 3
Forno 3
Carvão 41
Moinho 3
Moinho 4
Graficos
Alarmes
Mais Equi.
Ref. Equi.
Fig.Pequenas
Forno 3
Permutador e Filtro do Arrefecedor
Aceit. ALARMES
TRANSPORTE MATÉRIAS PRIMAS
Nascente
Poente
Mistura
Calcario Pobre
Calcario Rico
ABASTECIMENTO CLINQUER
activa
desactiva
Filler
Gesso
Gesso e Outros
CLK SL06
P/ Stock Polar
Aditivos
Gesso
Clinquer
N3
N4
Ar de Comando
Ar de Reserva
SL02
Automático
Disturbio
S3W04 - Peso Silo 3 Carvão
Unidades: Ton
Minimo: +0,0
Máximo: +99,6
Equipamento
Descrição Equipamento
Tipo Alarme
K3M128
Bomba Injecção Agua Torre (1)
Defeito Electrico
K3M133
Bomba Injecção Agua Torre (3)
Defeito Electrico


Forno 3
Filtro Mangas
TRANSPORTE MATÉRIAS PRIMAS
ABASTECIMENTO CLINQUER
Limpeza Mangas à Distância
Segurança Filtro
Adição de Amónia
Aceit. ALARMES
Equipamento
Descrição Equipamento
Tipo Alarme
K3M128
Bomba Injecção Agua Torre (1)
Defeito Electrico
K3M133
Bomba Injecção Agua Torre (3)
Defeito Electrico
S3W04 - Peso Silo 3 Carvão


TRANSPORTE MATÉRIAS PRIMAS
Calcario Rico
Calcario Pobre
Janela de Comandos / Informações
FUZZY K3 - UNIT 1
NOME
LIDA
ESCRITA
GRUPO
FUZZY L3 - UNIT 1
ARRANQUE
PARAGEM
Aditivos
Gesso
Clinquer

Tag	Variable	Unit	Description
L3AI01201.A1	L3AI01201	bar	Pressão de Ar da queima
L3M101.SUPERV	L3M101		Motor a entrada do forno
L3M112.SUPERV	L3M112		
L3M109.SUPERV	L3M109		
L3M106.SUPERV	L3M106		
L3I01.A1	L3I01	Amp	Intensidade no motor L3M106
L3AI01202.A1	TERMO	ºC	Temperaturanacamaratermica
L3T12.A1	L3T12	ºC	Temperatura no forno
L3S01_A0.A0	L3S01A0	%	Percentagem da velocidade de rotação do motor do forno
L3S01.A1	L3S01	rpm	Velocidade de rotação do motor do Forno
L3M330.SUPERV	L3M330		Motor à saída do forno
L3T305.A1	L3T305	ºC	Temperatura do material saída do forno
L3P311.A1	L3P311	mbar	Pressão à saída do forno
L3M335.SUPERV	L3M335		Motor do arrefecedor do forno
L3M405.SUPERV	L3M405		Motor de tapete de transporte de clinquer
L3M433.SUPERV	L3M433		Motor de transporte de stocks
L3M202.SUPERV	L3M202		
L3M209.SUPERV	L3M209		
L3M207.SUPERV	L3M207		
L3F200.A1	L3F200	t/h	Tonelada/hora a saída do forno
L3M3237.SUPERV	L3M237		Motor à saída do fornopuxa o material
L3W200.A1	L3W200	Ton	Toneladas
L3Q200.A1	L3Q200	%CO	Percentagem de oxido de carbono Final
L3CALTRCTR.CV	L3CALTRCTR	%	
L3CALTRCTR.SET	L3CALTRCTR	bar	
L3AI0972.A1	L3CALTR_P01	bar	
L3AI0970.A1	L3CALTR_Q01	I/h	
L3AI0971.A1	L3CALTR_S01	rpm	
L3AI0973.A1	L3CALTR_T01	ºC	
L3G01.A1	L3G01	%	
L3G01_OUT.A0	L3G01	%	
L3F03.A1	L3F03	I/h	Retorno
L3T324.A1	L3T324	ºC	Temperatura a entrada da ventilação
L3T322.A1	L3T322	ºC	Temperatura a dentro da ventilação
L3T322_OUT.A0	L3T322	ºC	Temperatura à saída da ventilação
L3Y363.SUPERV	L3Y363		Comutadora entrada ….
L3P318.A1	L3P318	mbar	Pressão no filtro de mangas
L3I312.A1	L3I312	Amp	Intensidade do motor L3M366
L3M366.SUPERV	L3M366		
L3S310.A1	L3S310	rpm	Rotaçõesporminuto de motor L3M366
L3T304.A1	L3T304	ºC	Temperaturana entrada do arrefecedor
L3I311.A1	L3I311	Amp	Intensidade do motor de arrefecimento L3M335

L3T326.A1	L3T326	ºC	Temperatura a saida do arrefecedor
L3T327.A1	L3T327		
L3T13.A1	L3T13	ºC	Temperatura de admissão do queimador
L3AI00063.A1	L3AI00063	Amp	Intensidade do motor L3M207
L3P02.A1	L3P02	mbar	Pressao de arnas entradas axial e radial do queimador
L3AI00060.A1	L3AI00060	kg/min	Axial
L3AI00062.A1	L3AI00062	kg/min	Radial
L3AI00061.A1	L3AI00061	kg/min	Central
L3P01.A1	L3P01	mbar	Pressao no queimador

AutomatoCarvão (S3)

Tag	Variable	Unit	Description
S3Q03.A1	S3Q03	ppm	Concentração de carvãonaatmosfera
S3W04.A1	S3W04	Ton	Toneladas de carvão no moinho
S3F04.A1	S3F04	t/h	Tonelada hora
	S3F04	%	Percentagem de tonelada hora
S3M832.SUPERV	S3M832		
S3M831.SUPERV	S3M831		
S3M831Y1.SUPERV	S3M831Y1		
S3M832Y1.SUPERV	S3M832Y1		
	S3F02		Debitocarvãoqueimaforno 3

Torre de Condicionamento (K3)

Tag	Variable	Unit	Description
K3L998.A1	K3L998	Nm3/h	Caudal a saída da chamine
K3L997.A1	K3L997	%	Dioxido de Carbono à saída (CO_2), pág. 9 PDF
K3L996.A1	K3L996	mg/m3	Dioxido de Nitrogenio (NO_2), pág. 9 PDF
K3L995.A1	K3L995	mg/m3	Acidocloridico
K3AI00212.A1	K3AI00212	mg/Nm3	Hafnium
K3AI00233.A1	K3AI00233	mg/m3	Particulas (2) / H2O; TOC; NH3; Temp.
K3AI00213.A1	K3AI00213	%	H2O
K3AI00230.A1	K3AI00230	mg/m3	Total de Carbonoorganico
K3AI00231.A1	K3AI00231	ºC	Temprature
K3Q05.A1	K3Q05	mg/m3	Particulas (1) O2; CO; NO; SO2
K3AI00313.A1	K3AI00313	%	Percentagem de oxigenio
K3AI00310.A1	K3AI00310	mg/m3	Oxido de Carbono (CO), pág. 9 PDF
K3AI00311.A1	K3AI00311	mg/m3	OxidoNitrico (NO), pág. 9 PDF
K3AI00312.A1	K3AI00312	mg/m3	Dioxido de Enxofre (SO2), pág. 9 PDF
K3P02.A1	K3P02	mbar	Pressão a entrada da torre de condicionamento
K3T16.A1	K3T16	ºC	Temperatura a entrada de torreconsicionamento
K3G12.A1	K3G12	%	Percentagem ….
K3G43.A1	K3G43	%	Entrada do 1P
K3T18.A1	K3T18	ºC	Temperatura a entrada do 1P
K3T20.A1	K3T20	ºC	Temperatura à saída do 1P
K3P13.A1	K3P13	mbar	Pressão à saída do 1P
K3G42.A1	K3G42	%	
K3T17.A1	K3T17	ºC	Temperatura a entrada 1N

K3T19.A1	K3T19	ºC	Temperatura a saida 1N
K3P12.A1	K3P12	mbar	Pressão a saida 1N
K3F07.A1	K3F07	t/h	Tonelada hora vinda do elevador para as entradas de 1N e 1P
K3T49.A1	K3T49	ºC	Temperatura a entrada 2P
K3T51.A1	K3T51	ºC	Temperatura a saida 2P
K3P15.A1	K3P15	mbar	Pressao a saida 2P
K3T46.A1	K3T46	ºC	Temperatura a entrada 2N
K3T50.A1	K3T50	ºC	Temperatura a saida 2N
K3P15B.A1	K3P15B	mbar	Pressao a saida 2N
K3Q03.A1	K3Q03	CO	Oxido de Carbono a entrada ciclone 3
K3Q02.A1	K3Q02	O2	Oxigenio a entrada ciclone 3
K3P03.A1	K3P03	mbar	Pressao a entrada do cilone 3
K3T21.A1	K3T21	ºC	Temperatura a entrada do ciclone 3
K3T52.A1	K3T52	ºC	Temperatura a saidaciclone 3
K3P14.A1	K3P14	mbar	Pressao a saida do ciclone 3
K3Y465.SET	K3Y465		
K3G14.A1	K3G14		
K3P04.A1	K3P04	mbar	Pressao a entrada 4N
K3T22.A1	K3T22	ºC	Temperatura a entrada 4N
K3T25.A1	K3T25	ºC	Temperatura a saida 4N
K3P05.A1	K3P05	mbar	Pressao a entrada 4P
K3T23.A1	K3T23	ºC	Temperatura a entrada 4P
K3T26.A1	K3T26	ºC	Temperatura a saida 4P
K3T53.A1	K3T53	ºC	Temperatura …..
K3P16.A1	K3P16	mbar	Pressao …..
K3T24.A1	K3T24	ºC	Temperatura a entrada do forno
K3P06.A1	K3P06	mbar	Pressao a entrada do forno
K3Q01.A1	K3Q01	O2	Injecção de oxigenio a entrada do forno
K3Q06.A1	K3Q06	CO	Injecção de Oxido de Carbono a entrada do forno
K3Q07.A1	K3Q07	NOx	Injecção de NOx a entrada do forno
K3AI00211.A1	K3AI00211	SO2	Injecção de SO2 a entrada do forno
K3M487.SUPERV	K3M487		
K3T16.A1	K3I16	Amp	Intensidade do motor K3M487
K3G37.A1	K3G37	%	Percentagem de abertura da valvulapara o ventilador K3M487
K3M489.SUPERV	K3M489		
K3F09.A1	K3F09	%	
K3P20.A1	K3P20	mbar	Pressao a saida do K3M489 0 K3M487
K3AI1150.A1	K3AI1150	ºC	Temperaturavinda do carvao 1
K3AI1151.A1	K3AI1151	ºC	Temperaturavinda do carvao 2
K3AI1152.A1	K3AI1152	ºC	Temperaturavinda do carvao 3
K3AI1153.A1	K3AI1153	ºC	Temperaturavinda do carvao 4
K3P19.A1	K3P19	mbar	Pressao no cilcone de carvao
K3T62.A1	K3T62	ºC	Temperaturadentro do ciclone de carvao
K3AI1161.A1	K3T62A	ºC	Temperatura a saida do ciclone de carvao
K3Q09.A1	K3Q09	CO	CO a saida do ciclone de carvao
K3Q08.A1	K3Q08	O2	O2 a saida do cilone de carvao

K3Q10.A1	K3Q10	Nox	NOx a saida do ciclone de carvao
K3M480C.SUPERV	K3M480C		Motor do despoeiramento
K3M461.SUPERV	K3M461		Motor que puxa o retorno gases
K3G36.A1	K3G36	%	Posição do registo do arterceario
K3T60.A1	K3T60	ºC	Temperatura do retorno
K3P17.A1	K3P17	mbar	Pressao do retorno
K3M404.SUPERV	K3M404		Ventilador para moinho de cru
K3I02.A1	K3I02	Amp	Intensidade do motor K3M404
K3I15.A1	K3I15	Amp	
	K3S02		velocidade do ventiladorDopol (saida da torre)
	K3F06		Debito da doseadoratremonhapoeiras
	K3G14		Posição do registodevisao material a saida
	K3F10		Velocidade do ventilador do filtro

ParametrosProcesso

Tag	Variable	Unit	Description
POLAB_ANA73	SiO2	%	Valor do resultado no POLAB (laboratorio)
POLAB_ANA74	Al203	%	Valor do resultado no POLAB (laboratorio)
POLAB_ANA75	Fe203	%	Valor do resultado no POLAB (laboratorio)
POLAB_ANA76	CaO	%	Valor do resultado no POLAB (laboratorio)
POLAB_ANA79	LSF	%	Valor do resultado no POLAB (laboratorio)
POLAB_ANA78	A/F	%	Valor do resultado no POLAB (laboratorio)
POLAB_ANA77	MS	%	Valor do resultado no POLAB (laboratorio)
POLAB_ANA81	R200	%	Valor do resultado no POLAB (laboratorio)
L3K3PCA0[0]	CAL LIVRE	%	Valor do resultado no POLAB (laboratorio)
OUTPUT_ANA73	SiO2	%	Valor do resultadovaloremserviço
OUTPUT_ANA76	CaO	%	Valor do resultadovaloremserviço
OUTPUT_ANA79	LSF	%	Valor do resultadovaloremserviço
OUTPUT_ANA77	MS	%	Valor do resultadovaloremserviço
OUTPUT_ANA81	R200	%	Valor do resultadovaloremserviço
CLIVRE_OUT	CAL LIVRE	%	Valor do resultadovaloremserviço
MANUAL_ANA73	SiO2	%	Valor do resultadovalor de entrada manual
MANUAL_ANA76	CaO	%	Valor do resultadovalor de entrada manual
MANUAL_ANA79	LSF	%	Valor do resultadovalor de entrada manual
MANUAL_ANA77	MS	%	Valor do resultadovalor de entrada manual
MANUAL_ANA81	R200	%	Valor do resultadovalor de entrada manual
K3CAL_LIVRE	CAL LIVRE	%	Valor do resultadovalor de entrada manual
K3RSD_ESP	Peso/Litro	g/L	Valor do resultadovalor de entrada manual
S3RES_R90	R90_Carvao	%	Valor do resultadovalor de entrada manual
FACTOR_CLK_K3	F. C. Clinquer		Valor do resultadovalor de entrada manual
FACTOR_TERMICO_L3	P. C. Carvão		Valor do resultadovalor de entrada manual
FACTOR_TERMICO_CA	P. C: Cmb. A.		Valor do resultadovalor de entrada manual
PESPC_CALTR	Peso E. Cmb. A		Valor do resultadovalor de entrada manual
CODIGO_LER_SET	CodigoCmb.		Valor do resultadovalor de entrada manual

Appendix B: The output of EventiC for 196 inputs over production rate (no. 137). In this example, the rate of data acquisition was 1 sample at every minute for a complete month.

1	0.99574	50	0.99574	99	0.99574	148	0.99574
2	0.9954	51	0.99574	100	0.99574	149	0.99574
3	0.99572	52	0.97414	101	0.99574	150	0.99574
4	0.99574	53	0.95198	102	0.99574	151	0.99612
5	0.98481	54	0.91663	103	0.99574	152	0.99574
6	0.98507	55	0.90881	104	0.99574	153	0.99574
7	0.35963	56	0.3994	105	0.9953	154	0.99574
8	0.37235	57	0.93579	106	0.99574	155	0.99553
9	0.9727	58	0.97435	107	0.99574	156	0.99486
10	0.97733	59	0.33091	108	0.9957	157	0.99574
11	0.99565	60	0.69423	109	0.99484	158	0.99574
12	0.2917	61	0.51453	110	0.99574	159	0.35288
13	0.95358	62	0.58751	111	0.99533	160	0.6773
14	0.89633	63	0.14386	112	0.99574	161	0.95798
15	0.79467	64	0.41667	113	0.99574	162	0.99574
16	0.98235	65	0.99574	114	0.9953	163	0.93881
17	0.91695	66	0.38884	115	0.99553	164	0.65642
18	0.95223	67	0.86137	116	0.99563	165	0.98049
19	0.98135	68	0.5784	117	0.9957	166	0.95535
20	0.99253	69	0.30277	118	0.99484	167	0.96653
21	0.99526	70	0.40633	119	0.9956	168	0.95309
22	0.99574	71	0.55416	120	0.99533	169	0.97184
23	0.32344	72	0.95512	121	0.99574	170	0.92823
24	0.99574	73	0.96012	122	0.99574	171	0.80935
25	0.979	74	0.7473	123	0.99574	172	0.68184
26	0.99574	75	0.99263	124	0.99556	173	0.98202
27	0.33428	76	0.99551	125	0.98195	174	0.99488
28	0.99574	77	0.99574	126	0.63893	175	0.94847
29	0.98879	78	0.9957	127	0.99467	176	0.97195
30	0.74788	79	0.99577	128	0.48112	177	0.92691
31	0.99307	80	0.9957	129	0.99474	178	0.99477
32	0.99574	81	0.9957	130	0.76579	179	0.99044
33	0.99398	82	0.99556	131	0.97851	180	0.99202
34	0.99574	83	0.99572	132	0.8	181	0.99574
35	0.99574	84	0.99574	133	0.95002	182	0.99574
36	0.99514	85	0.99533	134	0.97888	183	0.98593
37	0.99533	86	0.99428	135	0.98535	184	0.99574
38	0.98753	87	0.99574	136	0.57358	185	0.99574
39	0.97463	88	0.99574	137	1	186	0.99574
40	0.97547	89	0.99549	138	0.99574	187	0.99574
41	0.99574	90	0.99567	139	0.99574	188	0.99567
42	0.45777	91	0.99547	140	0.92449	189	0.99221
43	0.66586	92	0.39112	141	0.98719	190	0.99574
44	0.97377	93	0.99574	142	0.99019	191	0.99574
45	0.98526	94	0.99374	143	0.98521	192	0.99574
46	0.99558	95	0.9934	144	0.99574	193	0.99574
47	0.99574	96	0.99574	145	0.99344	194	0.49791
48	0.99409	97	0.99574	146	0.99574	195	0.99574
49	0.99574	98	0.99574	147	0.99219	196	0.99574

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1E+09	125	0.59	8078	2.2	2.2	-3.0525	0.1349	10.14	19.134	118.63	2.8975	366.98	221.29	299.14	10.114	763.77	759.81	1299.8	837.17	1113.4	1.9	0.4088	0	256.25	0	11.048	0	23812	0.1167
1E+09	109	0.59	8078	2.2	2.2	-3.0525	0.019	10.232	19.073	118.75	3.2126	360.83	214.85	280	10.366	792.34	755.84	1299.8	835.98	1116.4	1.9	0.4188	0	257.46	0	10.438	0	23812	0.125
1E+09	109	0.59	8078	2.2	2.2	-3.0525	0.019	10.232	19.093	118.75	3.975	360.27	214.85	280	10.366	781.63	767.74	1299.8	837.17	1116.8	1.9	0.1538	0	256.44	0	9.9191	0	23812	0.1417
1E+09	109	0.59	8078	2.2	2.2	-3.0525	0.0834	10.152	19.134	118.87	8.2544	363.07	211.39	291.71	10.313	778.45	767.74	1299.8	837.97	1117.6	1.9	0.2075	0	259.78	0	12.697	0	23812	0.1167
1E+09	109	0.59	8078	2.2	2.2	0	0.077	10.152	19.155	118.87	3.0703	363.07	211.39	291.71	10.313	774.49	768.14	1299.8	837.17	1122.6	1.9	0.25	0	259.81	0	12.971	0	23812	0.1458
1E+09	109	0.59	8078	2.2	2.2	-3.0525	0.0834	10.152	19.093	119	3.2126	363.07	211.39	291.71	10.305	773.69	765.36	1299.8	835.98	1136.5	1.9	0.3588	0	259.89	0	11.14	0	23812	0.15
1E+09	109	0.59	8078	2.2	2.2	-3.0525	-0.0068	10.232	19.401	119	5.7437	362.51	205.45	276	10.29	785.2	761.79	1299.8	838.76	1140.7	1.9	0.2913	0	255.86	0	9.9496	0	23812	0.1458
1E+09	109	0.59	8078	2.2	2.2	-3.0525	-0.0068	10.232	19.463	119	3.0805	363.07	205.45	276	10.29	782.02	762.58	1299.8	840.75	1144.8	1.9	0.2213	0	257.09	0	11.415	0	23812	0.1417
1E+09	109	0.59	8078	2.2	2.2	-6.105	0.0383	10.335	19.401	119	2.3893	358.04	206.94	286	10.091	784.4	761.79	1299.8	841.14	1144	1.9	0.1238	0	258.23	0	14.65	0	23812	0.1917
1E+09	109	0.59	8078	2.2	2.2	-3.0525	0.0319	10.335	19.422	119.12	3.2736	357.48	206.94	286	10.091	779.25	765.36	1299.8	841.14	1141.5	1.9	0.23	0	255.63	0	11.934	0	23812	0.1708
1E+09	109	0.59	8078	2.2	2.2	-3.0525	0.019	10.364	19.381	119.12	4.9508	356.36	216.83	262.86	10.068	774.09	766.55	1299.8	841.94	1132.3	1.9	0.3975	0	256.06	0	12.483	0	23812	0.1833
1E+09	109	0.59	8078	2.2	2.2	-3.0525	0.019	10.364	19.217	119.12	2.4604	356.92	216.83	262.86	10.068	751.47	771.71	1299.8	841.54	1126.8	1.9	0.38	0	259.25	0	12.544	0	23812	0.1833
1E+09	109	0.59	8078	2.2	2.2	-3.0525	-0.0325	10.41	19.196	119.12	2.3486	356.36	217.33	274	10.244	778.45	770.12	1299.8	839.95	1122.6	1.9	0.3463	0	259.79	0	11.781	0	23812	0.1875
1E+09	109	0.59	8078	2.2	2.2	0	-0.0261	10.41	19.217	119.12	2.3486	356.36	217.33	274	10.244	753.46	777.66	1299.8	840.75	1117.6	1.9	0.095	0	257.85	0	19.991	0	23812	0.2167
1E+09	109	0.59	8078	2.2	2.2	3.0525	0.0447	10.358	19.217	119.12	4.3003	357.48	211.39	271.14	10.236	777.26	767.74	1299.8	841.14	1118	1.9	0.1113	0	259.73	0	11.018	0	23812	0.2375
1E+09	109	0.59	8078	2.2	2.2	3.0525	0.0447	10.358	19.196	119.12	2.4198	357.48	211.39	271.14	10.236	756.23	778.06	1299.8	843.13	1111.8	1.9	0.1013	0	257.99	0	11.415	0	23812	0.2125
1E+09	109	0.59	8078	2.2	2.2	3.0525	-0.0261	10.399	19.196	119.36	2.4503	355.24	208.92	260	10.229	748.3	779.25	1299.8	842.73	1114.3	1.9	0.465	0	257.59	0	12.971	0	23812	0.2208
1E+09	109	0.59	8078	2.2	2.2	3.0525	-0.0261	10.399	19.196	119.24	2.3995	354.69	208.92	260	10.229	752.66	783.21	1299.8	841.14	1112.6	1.9	0.3575	0	255.63	0	13.887	0	23812	0.2375
1E+09	109	0.59	8078	2.2	2.2	3.0525	-0.0197	10.399	19.155	119.24	2.867	354.69	208.92	260	10.229	770.12	783.61	1299.8	840.75	1115.9	1.9	0.2813	0	257.99	0	12.727	0	23812	0.2458
1E+09	109	0.59	8078	1.8	1.8	3.0525	-0.1227	10.444	18.929	119.24	2.308	361.95	222.28	263.71	10.167	752.66	782.42	1299.8	839.16	1126.4	1.9	0.0688	0	255.01	0	11.293	0	23812	0.2583
1E+09	109	0.59	8078	1.8	1.8	3.0525	-0.1227	10.444	18.888	119.36	2.4299	361.95	222.28	263.71	10.16	752.66	787.98	1299.8	837.97	1118	1.9	0.1775	0	254.35	0	13.307	0	23812	0.2583
1E+09	109	0.59	8078	1.8	1.8	0	-0.0325	10.422	18.785	119.24	2.6637	355.24	221.29	260.29	10.06	767.74	795.51	1299.8	837.17	1102.1	1.9	0.1613	0	259.61	0	12.422	0	23812	0.2625
1E+09	109	0.59	8078	1.8	1.8	0	-0.0325	10.422	18.724	119.24	3.0703	355.24	221.29	260.29	10.06	793.93	782.42	1299.8	834.79	1100.9	1.9	-0.4063	0	259.6	0	6.1651	0	23812	0.2583
1E+09	109	0.59	8078	1.8	1.8	3.0525	-0.0647	10.358	18.6	119.24	2.4909	353.01	216.83	256.86	10.313	761.79	789.96	1299.8	833.6	1107.6	1.9	-0.4025	0	260.11	0	11.14	0	23812	0.2708
1E+09	109	0.59	8078	1.8	1.8	0	-0.0647	10.353	19.032	119.36	2.5418	353.01	216.83	256.86	10.313	770.52	782.42	1299.8	832.02	1108	1.9	-0.3725	0	260.28	0	12.544	0	23812	0.2833
1E+09	109	0.59	8078	1.8	1.8	3.0525	-0.0261	10.422	19.278	119.48	2.867	364.75	202.48	261.71	10.152	752.66	772.1	1299.8	832.02	1106.3	1.9	-0.3663	0	258.93	0	6.1651	0	23812	0.2917
1E+09	109	0.59	8078	1.8	1.8	3.0525	-0.0197	10.422	19.114	119.36	4.0563	365.3	202.48	261.71	10.152	765.76	779.25	1299.8	832.02	1105.5	1.9	-0.4175	0	257.99	0	11.506	0	23812	0.2958
1E+09	109	0.59	8078	1.8	1.8	6.105	0.0125	10.444	19.155	119.36	12.778	369.22	220.3	346.28	10.114	769.33	775.68	1299.8	831.22	1101.7	1.9	-0.2025	0	256.28	0	10.743	0	23812	0.2917
1E+09	109	0.59	8078	1.8	1.8	6.105	0.0125	10.444	19.052	119.24	3.8327	368.66	220.3	346.28	10.106	745.92	787.58	1299.8	831.22	1109.2	1.9	0.33	0	260.78	0	14.619	0	23812	0.3042
1E+09	109	0.59	8078	1.8	1.8	12.21	-0.0712	10.433	19.217	119.36	4.4426	382.07	214.36	359.7	10.075	750.28	789.96	1299.8	831.62	1112.2	1.9	0.225	0	255.9	0	13.337	0	23812	0.3208

31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
36.963	0	60.125	0	99.985	99.985	99.191	1698.2	658.67	657.04	-49.992	0.495	6.7571	15.399	464225	1	24	60	0	1	0	-47.379	-22.005	-14.165	-12.618	-1.0688	-38.968	-43.733	-22.909	-27.835
36.963	0	60.095	0	99.985	99.985	99.191	1698.2	661.6	659.48	-49.992	0.3658	8.2692	15.874	466595	1	24	60	0	1	0	-46.954	-21.625	-13.748	-12.368	-1.0356	-38.563	-43.305	-19.429	-27.102
36.967	0	60.034	0	99.985	99.985	99.191	1687.8	656.53	654.11	-49.992	0.3658	8.222	15.883	464407	1	24	60	0	1	0	-47.367	-22.073	-14.138	-12.4	-1.1081	-38.456	-43.84	-23.275	-27.359
36.963	0	60.125	0	99.985	99.985	99.191	1698.2	658.09	656.55	-49.992	0.2367	7.2769	15.874	461125	1	24	60	0	1	0	-46.938	-21.465	-14.22	-12.365	-1.1094	-39.951	-43.562	-20.784	-27.285
36.963	0	60.034	0	99.985	99.985	99.191	1694.1	657.31	655.33	-49.992	0.2582	7.2769	15.874	462948	1	24	60	0	1	0	-47.017	-21.59	-13.703	-12.225	-1.0088	-40.614	-43.605	-23.44	-27.578
36.963	0	60.125	0	99.985	99.985	99.191	1689.9	656.53	654.11	-49.992	0.2367	7.3242	15.874	463313	1	24	60	0	1	0	-46.813	-21.19	-13.875	-12.29	-1.0625	-40.208	-43.562	-21.92	-27.23
36.967	0	60.095	0	99.985	99.985	99.191	1696.8	656.72	654.11	-49.992	0.2798	6.4264	15.706	463495	1	24	60	0	1	0	-46.717	-21.823	-13.798	-12.5	-1.0638	-40.037	-43.199	-22.835	-26.864
36.963	0	60.034	0	99.985	99.985	99.191	1703	657.11	654.6	-49.992	0.2798	6.4264	15.706	466960	1	24	60	0	1	0	-46.675	-21.615	-13.618	-12.148	-1.0481	-39.139	-42.964	-21.956	-26.681
36.963	0	60.125	0	99.985	99.985	99.191	1696.8	660.43	657.78	-49.992	0.495	7.7022	15.846	467142	1	24	60	0	1	0	-46.842	-21.718	-13.738	-12.25	-0.9563	-39.951	-43.369	-21.627	-26.901
36.971	0	60.095	0	99.985	99.985	99.191	1698.2	656.33	654.11	-49.992	0.495	7.7022	15.846	462948	1	24	60	0	1	0	-46.904	-21.948	-13.888	-12.333	-0.8875	-40.421	-43.305	-21.828	-27.304
36.958	0	60.095	0	99.985	99.985	99.191	1698.2	654.57	652.16	-49.992	0.2152	7.6077	15.753	466413	1	24	60	0	1	0	-46.717	-21.915	-13.703	-12.698	-1.0219	-40.955	-43.348	-20.913	-27.963
36.963	0	60.034	0	99.985	99.985	99.191	1699.6	657.7	655.33	-49.992	0.2152	7.6077	15.753	466230	1	24	60	0	1	0	-46.838	-21.84	-14.123	-12.47	-1.0294	-39.844	-43.605	-24.3	-27.67
36.967	0	60.125	0	99.985	99.985	99.191	1694.1	656.53	654.36	-49.992	0.4089	7.7022	15.697	471336	1	24	60	0	1	0	-46.696	-21.745	-13.84	-12.25	-1.025	-39.353	-43.605	-24.447	-27.395
36.963	0	60.095	0	99.985	99.985	99.191	1694.1	657.31	655.58	-49.992	0.4089	7.655	15.697	462037	1	24	60	0	1	0	-46.683	-21.88	-13.83	-12.635	-1.0788	-39.673	-43.092	-22.78	-27.285
36.963	0	60.095	0	99.985	99.985	99.222	1696.8	659.07	656.07	-49.992	0.1937	6.6154	15.613	467325	1	24	60	0	1	0	-46.696	-21.57	-13.515	-12.105	-0.9381	-38.798	-43.455	-21.077	-27.487
36.967	0	60.034	0	99.985	99.985	99.191	1696.8	656.72	655.33	-49.992	0.1937	6.6154	15.613	459666	1	24	60	0	1	0	-46.663	-21.353	-13.558	-12.038	-0.8556	-38.263	-43.092	-18.074	-27.047
36.958	0	60.095	0	99.985	99.985	99.191	1696.8	656.14	653.62	-49.992	0.3013	7.7022	15.585	468054	1	24	60	0	1	0	-47.233	-21.918	-14.42	-12.368	-0.9338	-39.631	-43.946	-19.319	-27.56
36.963	0	60.095	0	99.985	99.985	99.191	1698.2	662.19	658.51	-49.992	0.3013	7.7022	15.585	463495	1	24	60	0	1	0	-47	-21.548	-14	-12.153	-0.8831	-38.712	-43.284	-23.202	-27.743
36.971	0	60.095	0	99.985	99.985	99.191	1705.1	659.07	657.04	-49.992	0.3013	7.7022	15.585	470242	1	24	60	0	1	0	-46.433	-21.163	-13.52	-12.253	-0.9125	-39.78	-42.964	-20.363	-27.304
36.963	0	60.034	0	99.985	99.985	99.191	1698.2	659.07	656.55	-49.992	0.3013	7.7967	15.622	468783	1	24	60	0	1	0	-46.679	-21.448	-13.78	-12.15	-0.9075	-40.186	-43.199	-25.344	-27.285
36.967	0	60.095	0	99.985	99.985	99.191	1698.2	658.09	655.58	-49.992	0.3013	7.7967	15.622	467142	1	24	60	0	1	0	-46.983	-21.738	-13.935	-12.128	-0.8963	-39.887	-43.818	-22.927	-27.871
36.967	0	60.125	0	99.985	99.985	99.191	1700.3	657.31	654.36	-49.992	0.3443	8.222	15.799	464407	1	24	60	0	1	0	-46.675	-21.415	-13.828	-12.333	-0.9269	-38.349	-43.391	-24.282	-27.084
36.963	0	60.095	0	99.985	99.985	99.191	1691.3	653.6	651.92	-49.992	0.3443	8.2692	15.809	461672	1	24	60	0	1	0	-46.85	-21.57	-13.778	-12.128	-0.9744	-40.101	-43.519	-22.909	-27.029
36.963	0	60.095	0	99.985	99.985	99.191	1701	657.11	655.58	-49.992	0.1937	8.3165	15.557	463313	1	24	60	0	1	0	-46.625	-21.26	-13.51	-12.258	-1.1188	-39.438	-43.263	-21.37	-27.322
36.963	0	60.095	0	99.985	99.985	99.222	1704.4	660.24	658.51	-49.992	0.1937	8.3165	15.557	464407	1	24	60	0	1	0	-47.458	-21.66	-13.913	-12.495	-1.2338	-40.079	-43.84	-21.59	-28.292
36.967	0	60.095	0	99.985	99.985	99.191	1707.2	661.6	659.48	-49.992	0.3013	6.4264	15.539	471701	1	24	60	0	1	0	-47.167	-21.868	-13.81	-12.628	-1.1025	-40.379	-43.882	-21.462	-28.146
36.963	0	60.034	0	99.985	99.985	99.191	1700.3	656.72	653.14	-49.992	0.2798	6.4264	15.539	470060	1	24	60	0	1	0	-46.954	-21.723	-13.758	-12.405	-1.2119	-38.776	-43.818	-22.286	-27.908
36.967	0	60.095	0	99.985	99.985	99.191	1706.5	657.11	654.36	-49.992	0.2582	8.033	15.66	464225	1	24	60	0	1	0	-46.858	-21.605	-13.693	-12.518	-1.2394	-38.157	-43.028	-21.645	-28.018
36.967	0	60.095	0	99.985	99.985	99.191	1698.2	657.5	654.6	-49.992	0.2582	8.0802	15.66	465683	1	24	60	0	1	0	-47.067	-21.683	-13.945	-12.365	-1.4106	-40.507	-43.733	-23.659	-27.78
36.967	0	60.095	0	99.985	99.985	99.191	1704.4	658.09	655.33	-49.992	0.3013	7.2769	15.809	465501	1	24	60	0	1	0	-47.35	-22.908	-14.255	-13.463	-1.1713	-39.118	-43.818	-22.048	-28.183

Appendix C: SCADA output for the kiln's 196 sensors in the first 30 samples (first 30 minutes)

61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90
-23.971	-9.147	-2.7316	-5.7897	49.901	-24.229	3.1931	0.1321	2.7914	-0.0004	0	0.4212	0.0012	1.2208	840.84	391.35	394.48	408.7	396.73	409.98	730.95	883.8	846.35	400.35	866.6	828.65	45.379	587.46	612.12	588.19
-23.476	-8.7899	-2.3653	-5.4845	49.626	-24.229	3.4106	0.1381	3.0815	-0.0004	-0.1667	0.3937	0.0012	1.2208	840.84	391.35	394.25	408.93	396.63	409.88	730.95	884	846.75	400.5	867.3	829.15	45.379	587.21	612.12	588.43
-25.362	-9.4583	-2.2066	-5.7531	49.992	-24.229	3.3006	0.1407	4.5509	-0.0005	0	0.4212	0.0012	1.2208	840.53	391.4	394.38	409.18	396.63	409.9	731.33	884.5	847.7	400.85	870.55	829.85	45.306	586.97	612.61	588.68
-23.623	-9.5529	-3.4885	-5.9118	49.809	-24.229	3.4156	0.1461	7.244	-0.0004	0	0.4349	0.0012	1.2208	840.84	391.68	394.4	409.48	396.73	409.9	731.59	885.85	848.55	401.05	871.4	830.3	45.306	587.21	612.12	588.92
-23.476	-9.9496	-2.1212	-6.1315	49.992	-24.229	3.4331	0.1464	2.9785	-0.0005	-0.1667	0.4395	0.0012	2.4416	840.53	391.45	394.4	409.55	396.73	409.88	731.66	886.05	848.75	401.15	871.05	830.45	45.379	587.21	612.85	589.41
-22.286	-9.6872	-2.9147	-5.9728	49.535	-24.229	3.12	0.1359	3.3197	-0.0005	-0.1667	0.467	0.0012	6.7145	840.84	391.88	394.35	410.18	396.78	409.93	732.08	886.95	849.95	401.45	877.35	831.65	45.416	586.97	613.34	588.92
-27.67	-9.1897	-1.291	-5.5211	49.809	-24.229	3.0606	0.1404	5.6376	-0.0005	0	0.5265	0.0012	9.7665	840.53	392	394.35	410.8	396.73	410.28	732.79	889.95	852.3	401.65	883.15	833.15	45.416	587.21	614.07	589.9
-25.948	-8.4603	-3.5251	-5.2648	49.901	-24.229	2.9875	0.13	2.7394	-0.0004	-0.1667	0.5356	0.0012	10.377	840.53	392.15	394.38	411.68	396.8	410.78	733.54	892.05	853.7	401.9	883.85	833.9	45.416	587.46	614.07	589.9
-22.451	-8.9944	-3.6228	-5.7043	49.809	-24.229	2.9994	0.1303	2.3141	-0.0004	-0.1667	0.5677	0.0012	1.2208	840.84	392.45	394	412.45	396.53	411.1	734.06	893.8	854.65	402.25	885.65	834	45.452	587.21	614.07	590.63
-23.623	-9.8031	-3.2199	-6.1804	49.626	-24.229	3.0625	0.1359	3.233	-0.0004	-0.1667	0.6272	0.0012	1.2208	840.53	392.65	394	412.73	396.55	411.28	734.48	894.65	854.85	402.35	885.45	833.75	45.416	587.21	614.07	590.87
-26.534	-9.5681	-1.6084	-5.8386	49.626	-24.229	3.0425	0.1358	4.5668	-0.0005	-0.1667	0.673	0.0012	1.2208	840.53	392.68	393.95	412.93	396.45	411.5	734.55	895.35	854.45	402.65	882.7	833	45.416	587.46	614.07	590.87
-25.564	-9.0768	-2.6705	-5.6676	49.626	-24.229	3.3256	0.1312	2.3309	-0.0005	-0.1667	0.6684	0.0012	1.2208	840.84	392.83	393.98	413.03	396.48	411.63	734.48	895.4	853.45	402.9	879.9	832.2	45.452	587.46	613.34	590.63
-23.916	-9.9985	-3.4397	-6.3269	49.809	-24.229	3.3325	0.1311	2.2789	-0.0005	-0.1667	0.7096	0.0012	1.2208	840.53	393	394.03	413.08	396.58	411.78	734.36	895.4	852.85	403.25	880.15	831.85	45.489	587.21	613.34	590.14
-24.85	-9.3667	-2.402	-5.7775	49.626	-24.229	3.2813	0.1354	2.3433	-0.0004	-0.1667	0.705	0.0012	1.2208	840.53	392.98	394.05	412.85	396.53	411.9	734.4	895.45	852.65	403.5	878.7	831.9	45.452	587.21	613.34	589.9
-24.575	-8.8082	-2.1822	-5.2281	49.809	-24.229	3.2963	0.1355	2.6107	-0.0005	0	0.7096	0.0012	1.2208	840.53	392.9	394.1	412.73	396.8	411.9	734.44	895.15	852.35	403.7	877.55	831.8	45.452	587.46	613.34	590.63
-21.444	-8.7166	-2.2188	-5.1061	49.626	-24.229	3.2088	0.1393	2.3266	-0.0005	-0.1667	0.6959	0.0012	1.2208	840.84	392.88	394.48	412.45	397.25	411.9	734.4	894.85	852	404	876.95	831.5	45.489	587.21	613.83	590.14
-23.605	-9.2019	-3.7571	-6.0827	49.626	-24.229	3.2169	0.1394	2.3789	-0.0004	-0.1667	0.6684	0.0012	1.2208	840.53	392.58	394.65	412.05	397.5	411.75	734.48	894.5	851.45	404.3	874.25	831	45.489	587.21	613.09	590.63
-25.93	-8.9089	-3.0856	-5.5333	49.901	-24.229	3.2525	0.1522	2.3291	-0.0005	-0.1667	0.673	0.0012	1.2208	840.53	392.03	394.75	411.15	397.55	411.63	733.88	892.8	849.85	404.7	871.05	830.25	45.489	587.21	613.09	590.63
-23.971	-8.6861	-2.4142	-5.338	49.809	-24.229	3.2575	0.1523	2.2077	-0.0004	-0.1667	0.6501	0.0012	1.2208	840.84	391.83	394.75	410.38	397.7	411.3	733.54	892.15	849.3	404.8	870.05	829.85	45.489	587.21	612.85	590.14
-23.568	-9.2599	-2.3898	-5.8141	49.626	-24.229	3.05	0.1501	2.2467	-0.0004	-0.1667	0.6501	0.0012	1.2208	840.53	391.6	395.63	408.58	399.13	410.58	733.5	891.2	849.95	405.2	873.05	830.8	45.562	587.46	614.07	589.65
-22.835	-9.2355	-3.403	-5.9606	49.626	-24.229	3.0563	0.1501	2.4201	-0.0004	-0.1667	0.618	0.0012	1.2208	840.53	391.65	397.18	406.88	401.18	409.43	733.58	891.35	850.75	405.5	873	830.85	45.489	587.21	615.54	591.12
-23.366	-9.0554	-3.1467	-5.8386	49.901	-24.229	3.1744	0.1554	2.5849	-0.0004	-0.1667	0.5951	0.0012	1.2208	840.84	391.63	397.88	405.28	402.25	408.73	733.54	890.5	849.6	405.85	868.8	830.05	45.599	587.7	615.05	591.12
-26.644	-8.5762	-3.0978	-5.2648	49.535	-24.229	3.4788	0.1576	3.1845	-0.0005	0	0.5539	0.0012	1.2208	840.84	391.65	398.5	403.83	403.15	407.98	733.01	888.55	848	406.2	865.55	828.95	45.599	587.46	615.29	589.9
-25.399	-9.0829	-2.7072	-5.4601	49.626	-24.229	3.1544	0.1608	2.4107	-0.0004	-0.1667	0.5585	0.0012	1.2208	840.53	391.7	398.73	403.05	403.48	407.73	732.9	887.95	847.75	406.4	867.8	829.35	45.636	587.21	615.05	590.14
-27.322	-9.8886	-2.9635	-6.0217	49.809	-24.229	3.1669	0.1611	2.4716	-0.0005	-0.1667	0.5723	0.0012	1.2208	840.84	391.55	398.9	402.15	403.65	407.13	732.71	888.05	848.4	406.7	871.4	830	45.599	587.21	614.56	589.9
-23.055	-9.9985	-2.6461	-6.1926	49.626	-24.229	3.095	0.1543	2.8201	-0.0005	0	0.5448	0.0012	1.2208	840.53	391.25	398.48	402.05	403.05	406.73	732.38	887.6	848.05	407.1	868.8	829.5	45.636	587.21	613.83	588.92
-24.483	-9.5742	-3.6838	-6.0217	49.809	-24.229	2.965	0.1706	3.8647	-0.0005	-0.1667	0.5585	0.0012	5.4937	840.84	391.08	397.35	402.73	401.83	406.78	732.11	887.05	847.9	407.5	868.55	829.65	45.636	586.97	612.85	588.43
-23.751	-9.0951	-3.0856	-5.5333	49.535	-24.229	2.9969	0.1745	10.042	-0.0004	-0.1667	0.4761	0.0012	10.377	840.53	390.73	396.05	403.95	400	406.98	731.7	886.15	847.6	407.8	867.1	830	45.672	587.21	612.12	588.43
-25.436	-9.9985	-4.0745	-6.4368	49.535	-24.229	2.9963	0.1737	3.7617	-0.0004	-0.1667	0.4761	0.0012	12.208	840.84	390.53	395.43	404.48	399.2	407.15	731.51	886.1	847.6	408	867.55	830.1	45.636	586.48	611.63	587.7
-29.098	-9.2446	-3.1344	-5.8386	49.992	-24.229	2.8038	0.1488	4.4093	-0.0004	-0.1667	0.5356	0.0012	1.2208	840.53	390.3	394.25	406.15	397.73	407.53	731.66	885.95	849	408.3	871.85	831.65	45.672	586.48	611.87	588.43

91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
719.98	534.11	880.45	1116.8	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
720.28	457.62	882.65	1120.6	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
720.59	484.97	884.85	1122.6	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
720.28	437.54	888.14	1121.8	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
720.59	530.69	888.51	1124.3	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
722.42	433.69	892.54	1153.6	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
723.94	518.72	896.2	1162.4	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
724.86	388.4	898.76	1162.4	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
724.55	404.64	900.96	1157.8	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
723.94	462.32	902.43	1152.4	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
723.64	476.42	904.26	1134.4	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
722.11	909.69	904.62	1121.8	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
721.5	438.4	904.26	1118.9	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
722.42	439.68	904.26	1109.7	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
721.81	578.97	903.16	1111.8	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
723.03	414.04	900.96	1099.2	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
720.89	484.97	898.03	1101.7	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
719.37	438.4	895.83	1100.9	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
718.45	840.04	893.27	1112.2	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
721.81	397.38	888.88	1139	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
722.42	545.22	883.38	1124.7	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
720.59	438.4	878.99	1092.9	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
719.37	503.77	874.23	1088.3	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
719.37	440.1	873.13	1100.9	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
720.28	754.59	871.3	1109.2	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
719.37	380.71	870.2	1100	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
718.76	541.37	870.2	1101.7	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
718.45	387.55	870.56	1095.4	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
717.84	399.51	871.3	1107.2	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51
720.28	429.42	872.03	1128.9	37	12.34	0	0	42.25	2.14	0	105.04	0	2.6	12.54	0	0	42.74	2.28	0	105.51	0	2.6	12.54	3.39	2.12	42.74	2.28	1.6	105.51

121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150
0	0	190205	2900	138	38.967	956.54	-0.8608	6.4093	1439.1	0	0	1.1964	11.995	25	-13.736	8.5627	5.8608	6552	378.82	69.815	223.26	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	138.81	36.099	958.85	-0.8608	6.3482	1514.2	0	0	1.1964	11.995	25	-13.736	8.6933	5.8608	6552	379.12	69.358	220.2	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	136.5	39.956	953.08	-0.7747	6.214	1561.5	0	0	1.1964	11.995	25	-18.315	8.6172	5.8608	6552	381.56	68.976	218.79	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	137.37	36.544	964.62	-0.8608	6.1804	1458.3	0	0	1.1964	11.995	25	-13.736	8.5807	5.8608	6552	379.43	68.213	220.2	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	137.19	43.17	954.23	-0.7747	6.1346	1399.7	0	0	1.2039	11.995	25	-18.315	8.5315	5.8608	6552	371.8	69.586	219.73	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	133.04	38.423	955.39	-0.7747	6.3818	1506.6	0	0	1.1964	11.995	25	-13.736	8.5564	5.8608	6552	373.63	69.51	224.2	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	135.58	37.582	964.62	-0.8608	6.5039	1419.3	0	0	1.1964	11.995	25	-13.736	8.629	5.8608	6552	379.43	69.052	220.2	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	137.54	36.445	958.85	-0.8608	6.5008	1411.6	0	0	1.2039	11.949	25	-13.736	8.5701	5.8608	6552	380.65	69.739	219.5	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	134.83	37.632	966.92	-0.8608	6.4276	1372	0	0	1.1964	11.995	25	-18.315	8.5363	5.8608	6552	377.29	71.494	220.2	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	135.12	37.731	964.62	-0.8608	6.4062	1403.4	0	0	1.1964	11.995	25	-13.736	8.5644	5.8608	6552	374.54	71.036	222.32	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	137.25	40.302	962.31	-0.8608	6.2719	1375.9	0	-0.4395	1.1964	11.995	25	-13.736	8.6287	5.8608	6552	377.6	69.052	220.91	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	138.81	39.758	958.85	-0.8608	6.1956	1474.5	0	-0.4395	1.1964	11.995	25	-18.315	8.5567	5.8608	6552	380.34	68.213	222.32	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	138.98	37.187	960	-0.8608	6.1621	1534	0	-0.4395	1.1964	11.949	25	-13.736	8.5659	5.8608	6552	382.78	69.892	224.9	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	134.31	34.517	960	-0.7747	6.1499	1656.1	0	-0.4395	1.1964	11.995	25	-18.315	8.6229	5.8608	6552	381.87	69.358	223.26	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	139.85	39.066	954.23	-0.8608	6.214	1561.5	0	0	1.1964	11.995	25	-18.315	8.5496	5.8608	6552	377.29	70.655	221.14	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	138	38.72	962.31	-0.8608	6.4673	1545.9	0	0	1.1964	11.995	25	-13.736	8.5655	5.8608	6552	375.76	68.747	222.08	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	138.17	39.363	962.31	-0.8608	6.5039	1573.4	0	0	1.2039	11.995	25	-18.315	8.5346	5.8608	6552	377.29	68.595	221.38	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	134.83	38.868	956.54	-0.8608	6.4459	1632.3	0	0	1.1964	11.995	25	-18.315	8.5513	5.8608	6552	380.34	69.892	223.49	2.2	1	1	1	4104	1	0	0
0	0	190205	2900	134.19	38.324	962.31	-0.8608	6.3543	1640.5	0	0	1.1964	11.995	25	-18.315	8.6194	5.8608	6552	381.87	69.434	222.55	1.8	1	1	1	4104	1	0	0
0	0	190205	2900	142.21	40.006	961.15	-0.8608	6.3208	1640.5	0	0	1.1964	11.949	25	-18.315	8.5278	5.8608	6552	379.43	67.984	224.2	1.8	1	1	1	4104	1	0	0
0	0	190205	2900	135	38.423	964.62	-0.8608	6.2597	1668	0	0	1.2039	11.995	25	-13.736	8.5672	5.8608	6552	374.85	68.747	223.26	1.8	1	1	1	4104	1	0	0
0	0	190205	2900	136.5	38.275	968.08	-0.8608	6.2536	1628.6	0	-0.4395	1.1964	11.995	25	-13.736	8.5468	5.8608	6552	376.98	69.739	224.2	1.8	1	1	1	4104	1	0	0
0	0	190205	2900	138.17	40.253	962.31	-0.8608	6.2475	1660.4	0	0	1.1964	11.995	25	-18.315	8.53	5.8608	6552	379.12	68.289	222.08	1.8	1	1	1	4104	1	0	0
0	0	190205	2900	138.98	37.83	956.54	-0.8608	6.4886	1628.6	0	0	1.1964	11.995	25	-13.736	8.5312	5.8608	6552	380.65	68.442	224.2	1.8	1	1	1	4104	1	0	0
0	0	190205	2900	135.69	38.918	964.62	-0.8608	6.5436	1628.6	0	0	1.2039	11.995	25	-18.315	8.5792	5.8608	6552	385.84	68.442	224.2	1.8	1	1	1	4104	1	0	0
0	0	190205	2900	136.44	36.841	961.15	-0.8608	6.452	1593.2	0	0	1.1964	11.995	25	-18.315	8.552	5.8608	6552	378.82	68.747	222.55	1.8	1	1	1	4104	1	0	0
0	0	190205	2900	137.25	39.61	958.85	-0.8608	6.3025	1545.9	0	0	1.1964	11.995	25	-13.736	8.5673	5.8608	6552	373.63	68.366	221.85	1.8	1	1	1	4104	1	0	0
0	0	190205	2900	136.67	38.571	964.62	-0.8608	6.2445	1553.6	0	0	1.1964	11.995	25	-13.736	8.5965	5.8608	6552	381.87	69.434	222.55	1.8	1	1	1	4104	1	0	0
0	0	190205	2900	130.44	38.819	964.62	-0.9469	6.2384	1739.1	0	0	1.1964	11.995	25	-13.736	8.5695	5.8608	6552	383.09	68.442	222.08	1.8	1	1	1	4104	1	0	0
0	0	190205	2900	136.5	37.632	962.31	-0.8608	6.2445	1498.6	0	0	1.1964	11.995	25	-18.315	8.602	5.8608	6552	384.01	69.129	221.14	1.8	1	1	1	4104	1	0	0

151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
1	1	1	1	1	1	2.575	-37.5	-0.5301	4.6238	11232	-0.1242	399.94	399.94	766.37	814.16	583.43	583.98	551.02	582.53	583.43	568.32	2.8105	2817.4	540.52	15.923	1349.4	150	982.33	84.603
1	1	1	1	1	1	2.575	-37.5	-0.5621	4.4026	11112	-0.125	399.94	399.94	762.25	809.22	574.64	593.59	535.36	563.23	580.96	574.64	2.8048	2817.4	530.04	15.923	1281.6	150	990.87	85.518
1	1	1	1	1	1	2.575	-37.5	-0.4836	4.1966	10928	-0.125	399.94	399.94	767.47	809.49	560.63	573.81	543.87	566.59	574.64	589.2	2.8161	2817.4	516.38	15.895	1351.8	150	993.01	86.068
1	1	1	1	1	1	2.575	-37.5	-0.5124	4.311	11072	-0.1242	399.94	399.94	769.66	813.61	574.36	573.26	545.25	582.95	577.38	524.37	2.8076	2817.4	516.38	15.886	1336.7	150	997.28	85.793
1	1	1	1	1	1	2.575	-37.5	-0.4837	4.1584	10928	-0.1241	399.94	399.94	763.35	816.91	580.68	576.01	552.11	580.85	579.31	571.89	2.8091	2817.4	516.38	15.923	1333.7	150	1002.4	86.525
1	1	1	1	1	1	2.575	-37.5	-0.5314	4.1584	10944	-0.1241	399.94	399.94	750.44	817.18	571.89	599.36	546.35	570.36	578.21	564.75	2.8076	2817.4	516.84	15.886	1322.4	150	1010.5	86.251
1	1	1	1	1	1	2.575	-37.5	-0.5563	4.166	10872	-0.125	399.94	399.94	766.64	817.46	581.23	565.02	546.62	555.69	574.09	551.29	2.8076	2817.4	508.19	15.923	1328	150	1014.4	87.075
1	1	1	1	1	1	2.575	-37.5	-0.4618	4.1126	10960	-0.125	399.94	399.94	772.96	822.68	575.74	565.57	558.98	597.62	590.57	562.28	2.8203	2817.4	508.64	15.84	1337.5	150	1020.8	86.708
1	1	1	1	1	1	2.575	-37.5	-0.4994	4.0592	10824	-0.1246	399.94	399.94	770.76	830.92	594.42	578.21	552.11	563.65	604.85	574.64	2.8091	2817.4	508.19	15.749	1326.2	150	1023.8	87.624
1	1	1	1	1	1	2.575	-37.5	-0.4739	3.9448	10888	-0.1245	399.94	399.94	773.78	827.35	576.56	572.99	552.39	565.33	594.42	605.4	2.8048	2817.4	508.64	15.602	1336.9	150	1024.2	86.708
1	1	1	1	1	1	2.575	-37.5	-0.5061	4.2118	11048	-0.1244	399.94	399.94	766.37	810.59	558.98	572.72	550.74	577.07	595.79	582.05	2.8048	2817.4	508.19	15.758	1336.9	150	1019.1	86.983
1	1	1	1	1	1	2.575	-37.5	-0.4354	4.3186	10904	-0.1246	399.94	399.94	789.44	830.92	562.83	565.57	532.61	569.94	569.97	621.33	2.8147	2817.4	508.64	15.675	1325.4	150	1012.2	87.532
1	1	1	1	1	1	2.575	-37.5	-0.4696	4.4636	11240	-0.1242	399.94	399.94	773.23	815.81	585.08	566.95	559.53	595.95	566.12	516.41	2.8091	2817.4	530.95	15.758	1344.7	150	1005.8	87.807
1	1	1	1	1	1	2.575	-37.5	-0.5967	4.4712	11224	-0.1241	399.94	399.94	783.4	827.07	580.13	565.02	554.04	583.36	563.38	587.82	2.8245	2817.4	528.22	15.923	1352.4	150	1001.6	88.448
1	1	1	1	1	1	2.575	-37.5	-0.5536	4.2805	11128	-0.1244	399.94	399.94	779	827.9	582.05	571.07	553.76	567.85	583.98	570.52	2.8105	2817.4	517.75	15.794	1351.2	150	995.58	86.617
1	1	1	1	1	1	2.575	-37.5	-0.4459	4.3797	11112	-0.1241	399.94	399.94	777.9	822.4	582.33	573.81	552.39	575.82	595.79	592.49	2.8091	2817.4	520.94	15.602	1358.2	150	984.89	87.166
1	1	1	1	1	1	2.575	-37.5	-0.4679	4.3186	11104	-0.1242	399.94	399.94	771.04	825.15	591.67	583.7	561.45	575.82	583.7	572.44	2.8119	2817.4	519.57	16.06	1352.3	150	980.19	86.251
1	1	1	1	1	1	2.575	-37.5	-0.3776	4.2347	10952	-0.125	399.94	399.94	763.62	826.8	580.68	566.4	565.02	567.43	608.15	583.7	2.8119	2817.4	529.13	15.602	1277.1	150	973.78	86.617
1	1	1	1	1	1	2.575	-37.5	-0.4744	4.5857	11312	-0.1248	399.94	399.94	774.33	825.42	574.36	579.31	551.02	574.98	569.14	561.45	2.8147	2817.4	530.95	15.675	1366.8	150	960.11	85.61
1	1	1	1	1	1	2.575	-37.5	-0.4561	4.6849	11480	-0.1241	399.94	399.94	779.83	836.96	581.51	584.8	552.11	574.14	593.59	619.41	2.8273	2817.4	540.52	15.84	1359.9	150	955.41	87.075
1	1	1	1	1	1	2.575	-37.5	-0.4642	4.5781	11320	-0.125	399.94	399.94	781.2	835.31	572.44	583.98	557.33	605.17	588.92	630.95	2.8119	2817.4	537.79	15.84	1390.1	150	947.29	86.159
1	1	1	1	1	1	2.575	-37.5	-0.5592	4.5323	11328	-0.1242	399.94	399.94	784.5	843.28	581.51	580.13	553.76	574.56	590.84	538.65	2.8105	2817.4	539.15	15.923	1250.6	150	944.3	85.243
1	1	1	1	1	1	2.575	-37.5	-0.3456	4.5933	11328	-0.125	399.94	399.94	784.5	832.02	564.47	572.44	547.72	563.23	567.5	604.03	2.8062	2817.4	537.79	15.84	1397.5	150	936.18	83.504
1	1	1	1	1	1	2.575	-37.5	-0.539	4.6086	11400	-0.1243	399.94	399.94	752.36	833.39	591.67	583.98	559.26	594.69	580.68	565.57	2.8076	2817.4	541.88	15.886	1353.2	150	933.62	83.321
1	1	1	1	1	1	2.575	-37.5	-0.6006	4.6467	11408	-0.1241	399.94	399.94	757.03	824.6	592.49	570.52	558.16	576.65	580.96	560.35	2.8203	2817.4	541.88	15.749	1366.8	150	934.47	83.504
1	1	1	1	1	1	2.575	-37.5	-0.4046	4.4712	11424	-0.1246	399.94	399.94	769.66	819.66	571.89	575.74	548.54	581.27	578.48	585.63	2.8175	2817.4	539.15	15.84	1337.3	150	940.03	83.412
1	1	1	1	1	1	2.575	-37.5	-0.4845	4.5857	11184	-0.1247	399.94	399.94	768.84	812.79	565.57	572.99	538.11	577.91	564.2	560.9	2.8062	2817.4	532.32	15.886	1337.9	150	940.03	82.13
1	1	1	1	1	1	2.575	-37.5	-0.5131	4.456	11232	-0.1241	399.94	399.94	756.48	816.36	575.74	572.72	539.48	557.78	586.45	585.63	2.8076	2817.4	531.41	15.886	1346.2	150	945.16	83.229
1	1	1	1	1	1	2.575	-37.5	-0.5141	4.5704	11264	-0.125	399.94	399.94	758.68	790.26	571.62	564.47	543.6	566.59	570.52	589.2	2.8105	2817.4	515.02	15.794	1366	150	957.12	82.039
1	1	1	1	1	1	2.575	-37.5	-0.3318	4.5399	11120	-0.1246	399.94	399.94	755.11	797.96	568.87	559.53	545.8	567.85	566.95	588.92	2.8161	2817.4	528.22	15.84	1351.2	150	963.1	81.489

181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196
13979	0	305.81	35.984	57.546	46.007	-32768	0.97	6	1	16384	0	-32768	0	3	47.543
13979	0	305.81	35.892	57.684	46.059	-32768	0.97	6	1	16384	0	-32768	0	3	47.421
13979	0	306.91	35.709	57.729	46.233	-32768	0.97	6	1	16384	0	-32768	0	3	47.269
13979	0	309.66	35.297	57.775	46.493	-32768	0.97	6	1	16384	0	-32768	-1.526	3	47.148
13979	0	312.41	35.205	57.684	46.566	-32768	0.97	6	1	16384	0	-32768	0	3	47.026
13979	0	315.89	35.068	57.729	46.672	-32768	0.97	6	1	16384	0	-32768	0	3	47.026
13979	0	323.39	35.068	57.729	46.94	-32768	0.97	6	1	16384	0	-32768	0	3	46.874
13979	0	331.63	35.068	57.775	47.062	-32768	0.97	6	1	16384	0	-32768	0	3	46.813
13979	0	334.38	34.976	57.684	47.218	-32768	0.97	6	1	16384	0	-32768	0	3	46.692
13979	0	335.3	35.022	57.684	47.343	-32768	0.97	6	1	16384	0	-32768	0	3	46.631
13979	0	337.13	35.022	57.638	47.493	-32768	0.97	6	1	16384	0	-32768	0	3	46.54
13979	0	338.23	35.022	57.546	47.699	-32768	0.97	6	1	16384	0	-32768	0	3	46.418
13979	0	336.21	34.885	57.638	47.905	-32768	0.97	6	1	16384	0	-32768	0	3	46.357
13979	0	332.73	34.839	57.684	47.972	-32768	0.97	6	1	16384	0	-32768	0	3	46.236
13979	0	332.37	34.839	57.638	48.031	-32768	0.97	6	1	16384	0	-32768	0	3	46.175
13979	0	330.9	34.747	57.729	48.348	-32768	0.97	6	1	16384	0	-32768	1.526	3	46.084
13979	0	328.52	34.61	57.684	48.497	-32768	0.97	6	1	16384	0	-32768	0	3	45.932
13979	0	325.77	34.61	57.638	48.746	-32768	0.97	6	1	16384	0	-32768	3.052	3	45.871
13979	0	322.3	34.656	57.638	48.953	-32768	0.97	6	1	16384	0	-32768	0	3	45.719
13979	0	320.46	34.61	57.729	49.081	-32768	0.97	6	1	16384	0	-32768	1.526	3	45.597
13979	0	316.25	34.335	57.775	49.199	-32768	0.97	6	1	16384	0	-32768	0	3	45.567
13979	0	311.67	34.152	58.004	49.43	-32768	0.97	6	1	16384	0	-32768	0	3	45.476
13979	0	307.65	34.107	58.004	49.595	-32768	0.97	6	1	16384	0	-32768	0	3	45.324
13979	0	304.72	34.061	57.958	49.83	-32768	0.97	6	1	16384	0	-32768	1.526	3	45.263
13979	0	302.88	34.061	57.958	49.953	-32768	0.97	6	1	16384	0	-32768	3.052	3	45.142
13979	0	299.77	34.107	58.096	50.067	-32768	0.97	6	1	16384	0	-32768	0	3	45.111
13979	0	296.84	34.335	57.958	50.289	-32768	0.97	6	1	16384	0	-32768	3.052	3	44.99
13979	0	293.36	34.427	57.958	50.481	-32768	0.97	6	1	16384	0	-32768	0	3	44.868
13979	0	290.98	34.564	58.004	50.627	-32768	0.97	6	1	16384	0	-32768	3.052	3	44.716
13979	0	290.07	34.564	57.912	50.743	-32768	0.97	6	1	16384	0	-32768	1.526	3	44.655