Plant-Wide Diagnosis: Cause-and-Effect Analysis Using Process Connectivity and Directionality Information

By

IYUN, Oluwatope Ebenezer

A Thesis submitted for the degree of
Doctor of Philosophy

October 2011

Centre for Process Systems Engineering
Department of Chemical Engineering and Chemical Technology
Imperial College of Science, Technology and Medicine
London SW7 2AZ

Table of Contents

Abstract

Production plants used in modern process industry must produce products that meet stringent environmental, quality and profitability constraints. In such integrated plants, non-linearity and strong process dynamic interactions among process units complicate root-cause diagnosis of plant-wide disturbances because disturbances may propagate to units at some distance away from the primary source of the upset. Similarly, implemented advanced process control strategies, backup and recovery systems, use of recycle streams and heat integration may hamper detection and diagnostic efforts.

It is important to track down the root-cause of a plant-wide disturbance because once corrective action is taken at the source, secondary propagated effects can be quickly eliminated with minimum effort and reduced down time with the resultant positive impact on process efficiency, productivity and profitability.

In order to diagnose the root-cause of disturbances that manifest plant-wide, it is crucial to incorporate and utilize knowledge about the overall process topology or interrelated physical structure of the plant, such as is contained in Piping and Instrumentation Diagrams (P&IDs). Traditionally, process control engineers have intuitively referred to the physical structure of the plant by visual inspection and manual tracing of fault propagation paths within the process structures, such as the process drawings on printed P&IDs, in order to make logical conclusions based on the results from data-driven analysis. This manual approach, however, is prone to various sources of errors and can quickly become complicated in real processes.

The aim of this thesis, therefore, is to establish innovative techniques for the electronic capture and manipulation of process schematic information from large plants such as refineries in order to provide an automated means of diagnosing plant-wide performance problems. This report also describes the design and implementation of a computer application program that integrates: (i) process connectivity and directionality information from intelligent P&IDs (ii) results from data-driven cause-and-effect analysis of process measurements and (iii) process know-how to aid process control engineers and plant operators gain process insight.

This work explored process intelligent P&IDs, created with AVEVA® P&ID, a Computer Aided Design (CAD) tool, and exported as an ISO 15926 compliant platform and vendor independent text-based XML description of the plant. The XML output was processed by a software tool developed in Microsoft® .NET environment in this research project to computationally generate connectivity matrix that shows plant items and their connections. The connectivity matrix produced can be exported to Excel® spreadsheet application as a basis for other application and has served as precursor to other research work. The final version of the developed software tool links statistical results of cause-and-effect analysis of process data with the connectivity matrix to simplify and gain insights into the cause and effect analysis using the connectivity information. Process knowhow and understanding is incorporated to generate logical conclusions.

The thesis presents a case study in an atmospheric crude heating unit as an illustrative example to drive home key concepts and also describes an industrial case study involving refinery operations. In the industrial case study, in addition to confirming the root-cause candidate, the developed software tool was set the task to determine the physical sequence of fault propagation path within the plant.

This was then compared with the hypothesis about disturbance propagation sequence generated by pure data-driven method. The results show a high degree of overlap which helps to validate statistical data-driven technique and easily identify any spurious results from the data-driven multivariable analysis. This significantly increase control engineers confidence in data-driven method being used for root-cause diagnosis.

The thesis concludes with a discussion of the approach and presents ideas for further development of the methods.

Declaration

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

Oluwatope Ebenezer IYUN

October 2011

Acknowledgements

Unless the Lord builds the house, those who build it labour in vain. Unless the Lord watches over the city, the watchman stays awake in vain (Psalm 127[1]). I give all glory and adoration back to the Lord Almighty for giving me the grace to complete this project. Only him is worthy of my ultimate praise.

This thesis would not have been possible without the guidance, advice, suggestions, constructive criticism, collaboration and the help of several individuals and organisations who in one way or another contributed and extended their valuable expertise in the preparation and completion of this study.

First and foremost, my utmost and sincere gratitude to my supervisor, Professor Nina F. Thornhill, ABB/RAEng chair of process automation for her continuous support of my research project, for her patience, guidance, understanding, pastoral care, motivation, innovative suggestions, encouragement, and expert advice. Her wealth of experience and attention to details are instrumental to the successful completion of my research project and production of this thesis.

Besides my supervisor, I would also like to thank Professor Nilay Shah, my collaborative academic for his interest in the PhD work.

I would like to acknowledge the support from my industrial sponsor, BP oil international for providing financial assistance and on-site practical training. My gratitude goes to my industrial supervisors: Engr. Natarajan Lakshman, Dr. Rajamani Murali, Dr. Rawi Zaid, Dr. Alleyne Ian, and Dr. Chopra Manan for their expert advice, insightful comments, encouragement, questions, and suggestions during our numerous meetings, phone calls, teleconferences and while I was on-site.

Input from collaborating industrial and software firms as well as their representatives to the research work is appreciated. I would like to thank David Birkett of AVEVA for efficiently resolving issues with the AVEVA software tool and John de Brugha of Hazid Technologies for sharing his experience in using intelligent P&IDs.

I would also like to thank the MSc students: Alabi David, Di Geronimo Gil Giovanni, and Dokmaithes Jittinun at Imperial College London for providing constructive feedback on their experience in using the software tool developed in this research project. Their users' experience, testing and evaluation reports helped in the implementation of improved functionalities in the final version of the software tool.

During the course of my studies at Imperial College, I have come across many interesting persons and made life-long friends, especially my fellow researchers (former and current) in the process automation group and in the research office. I will miss the company of Thambirajah Jegatheeswaran (Jerry), Ikram Waqas, Sharifzadeh Mahdi, Chen Su Liang (Nick), Cecílio Inês, Patel Mayank (May), Hosseini Seyed (Ali), Lee Chang-gun (Collins), Dunnett Alex, Xenos Dionysios, Sara Budinis, Erlend Meland, Jørgen Spjøtvold, and Sebastian Beez.

I would like to thank the staff of chemical engineering department at Imperial College London for providing the needed administrative support. I would not forget Susi Underwood in a hurry for her incessant emails and reminders in ensuring that deadlines are met.

Last but not the least, my sincere and heartfelt appreciation goes to my family for enduring my long absences away from home during the course of the PhD work. I would like to start with my darling and loving wife, Modupe Iyun (Dupsy) who has been a source of inspiration and encouragement to me before, during and after the research program. I am also indebted to my sons: Richard, Kelvin and Nelson for putting up with dad's incessant absences from home throughout the course of the PhD project. I promise to make it up to you guys. I would also like to thank my mom, brothers, sister, brothers-in-law, and sisters-in-law for their sustained prayers and encouragement. I would not forget the companionship provided by Rotimi (Tunde) Adegbite, Kayode (Jacob) Oguntimilehin, and Jimmy (Olu) Ashana during my stay in the city of London.

IYUN, O. E.

Imperial College London
August, 2011

List of Figures

List of Tables

Abbreviations

Following are the lists of abbreviations used in this report and their corresponding meanings. Every attempt has been made to define each abbreviation where it is first used in the report to aid understanding of the material.

Acronym	Definition
AAE	Average Absolute Error
ACF	Auto Covariance Function
AI	Artificial Intelligence
ANN	Artificial Neural Networks
APC	Advanced Process Control
API	Application Programming Interface
ARMA	Auto Regressive Moving Average
BG	Bond Graph
BPCS	Basic Process Control System
CAE	Computer Aided Engineering
CAEX	Computer Aided Engineering eXchange
CAD	Computer Aided Drawing
CR	Compensatory Response
CVSS	Canonical Variate State Space
DBMS	Database Management System
DCS	Distributed Control System
DTD	Document Type Definition
EKF	Extended Kalman Filtering
ERP	Enterprise Resource Planning
ESDG	Extended Signed Directed Graph
FDD	Fault Detection and Diagnosis
FDI	Fault Detection and Isolation
FFT	Fast Fourier Transform
FT	Fourier Transform
GUI	Graphical User Interface
IAE	Integral Absolute Error
ICA	Independent Component Analysis

Acronym	Definition
IEC	International Electrotechnical Commission
iP&ID	Intelligent Piping and Instrumentation Diagram
IR	Inverse Response
ISO	International Organization for Standardization
JVM	Java Virtual Machine
KBS	Knowledge Based System
KPI	Key Performance Indices/Indicators
OSD	Output Standard Deviation
ISA	Instrumentation, Systems and Automation
LISP	List Processing
LINQ	Language Integrated Query
PID	Proportional, Integral and Derivative
P&ID	Piping and Instrumentation Diagram
SDG	Signed Directed Graphs
SISO	Single-Input-Single-Output
MIMO	Multi-Input-Multi-Output
MPC	Model Predictive Control
MSPM	Multivariate Statistical Process Monitoring
MVPC	Multivariable Predictive Control
NMF	Non-Negative Matrix Factorisation
O-O	Object Oriented
PAS	Publicly Available Specification
P&ID	Piping and Instrumentation Diagram
PCA	Principal Component Analysis
PDA	Plant Disturbance Analyser
PFD	Process Flow Diagram
PLS	Partial Least Square
RDL	Reference Data Library
RPCA	Recursive Principal Component Analysis
SCM	Supply Chain Management
SDG	Signed Directed Graph
SP	Set Point
TE	Tennessee Eastman
UI	User Interface
URI	Uniform Resource Identifier

Acronym	Definition
QTA	Qualitative Trend Analysis
NN	Neural Network
WPF	Windows Presentation Foundation
XAML	eXtensible Application Mark-up Language
XBAP	XAML Browser Application
XML	eXtensible Mark-up Language

1 Introduction

This thesis presents findings from the research conducted into establishing innovative ways to capture and manipulate information from a process schematic in order to give an automated means of diagnosing plant-wide performance problems and performing process cause and effect analysis. The core objectives of the research work presented were to:

- Automate the process of capturing electronic connectivity information in large plants such as petroleum refineries.
- Manipulate connectivity information computationally to generate the connectivity matrix.
- Link connectivity information with results from data-driven cause-and-effect analysis of the process measurements.
- Incorporate process know-how to draw logical conclusions about the causes of disturbances.
- Produce a fully documented and tested software tool with good graphical user interface to demonstrate practical application of research findings. The software tool is a key deliverable required by the project sponsor.
- Validate research outputs with case studies.

Connectivity information refers to a specification of items in the plant and the connections between them in the form that can be manipulated algorithmically. An example of process know-how is the existence and mechanism of the destabilization that heat integration can have on a process. According to Thornhill & Horch (2007), a plant-wide process diagnosis approach implies that the distribution of a disturbance is mapped out across the plant, and the location and nature of the cause of the disturbance are determined with a high probability of being right the first time. Oscillations in process variables are a common form of plant-wide disturbance (Choudhury, *et al*., 2008).

The remaining sections of Chapter one are organized as follows. Section 1.1 introduces the problem addressed in the thesis. It provides background information necessary for the understanding of the origin and nature of the problem. Section 1.2 discusses the motivation behind the research work while section 1.3 enumerates specific outputs and contributions from the research work. Sections 1.4 and 1.5 discuss the research scope and the approach adopted to accomplish the research objectives. Section 1.6 provides a roadmap for the remaining part of the thesis. The chapter ends with a summary of the material discussed.

1.1 Background to the Problem

One of the fundamental objectives of process control is to transfer variability away from key process variables to less critical variables where such variability can be accommodated such as a buffer tank (Luyben, *et al*., 1999). However, due to the high level of interactions among plant units and the highly-coupled design nature of modern chemical process plants optimised to utilise recycle streams, heat integration and operate with reduced inventory, disturbances, such as oscillations, originating from a localized source simply propagate and manifest in other units within the plant as secondary upsets (Thornhill and Horch, 2007). Academic and industrial solutions have focused mostly on an individual control loop or equipment unit.

1.1.1 Single-Input Single-Output (SISO) Process Unit Control Strategy

The illustrations presented in this section are based on or taken from the data-driven root-cause diagnosis work reported by Thornhill, *et al*., (2003). Using data-driven analysis and knowledge about the process structure, the authors found a sticking valve as the root-cause of the plant-wide oscillations. Figure 1 is a SISO process unit taken from the process in Figure 2. When considered as an isolated unit, any deviation (variability) from the decanter's fluid level and its set point (reference) will be corrected by the level controller LC2 by adjusting the control valve. The plots on the right hand side of Figure 1 show the controller output, LC2op and the proxy flow measurement through the valve. The plots show that both LC2op and flow measurements exhibit oscillatory behaviour.

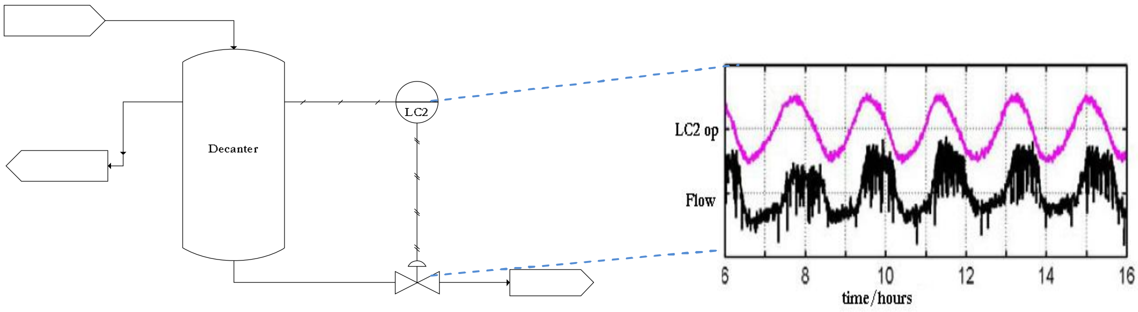


Figure 1: Level control in an isolated single-input single-output process unit. Adapted from Thornhill, *et al*., (2003)

However, since the decanter unit is an integral part of the overall process plant, actions taken by LC2 in order to maintain the correct fluid level in the decanter do propagate and affect other units within the plant as evidenced by process variables time trends in Figure 3. The locations of the disturbed tags are identified as black dots placed by hand in Figure 2. A tag is the name given to a measurement of calculated variable (such as controller output) that is recorded in the control system. The SISO decanter unit shown in Figure 1 is marked with a red hexagon on Figure 2.

Research and development in techniques for diagnosing and improving SISO control loop performance assessment and benchmarking is mature and well established in process industries (Harris, *et al*., 1999; Jelali, 2006; Qin, 1998; Yu, *et al*., 2010). This control and performance assessment mechanisms considers individual control loop and thus assumes that the units downstream of the unit under control remain unaffected by the controller action.

1.1.2 Plant-Wide Approach to Disturbance Detection and Diagnosis

Thornhill, *et al*., (2003) analysed routine data of process in Figure 2 and found LC2 (tag 22) as the root cause of the plant-wide disturbance. By plotting the valve input in LC2 control loop and the proxy flow F13 (tag 29) through the valve, the authors found LC2 control loop valve to be sticking, causing persistent limit cycle oscillation across the plant. Knowledge about the process layout was utilized in reaching the conclusion that tag 29 marked with a green hexagon in Figure 2 is a proxy measurement for the flow through tag 22 marked with a purple hexagon in Figure 2. The time trends of measured process tags are shown in Figure 3.

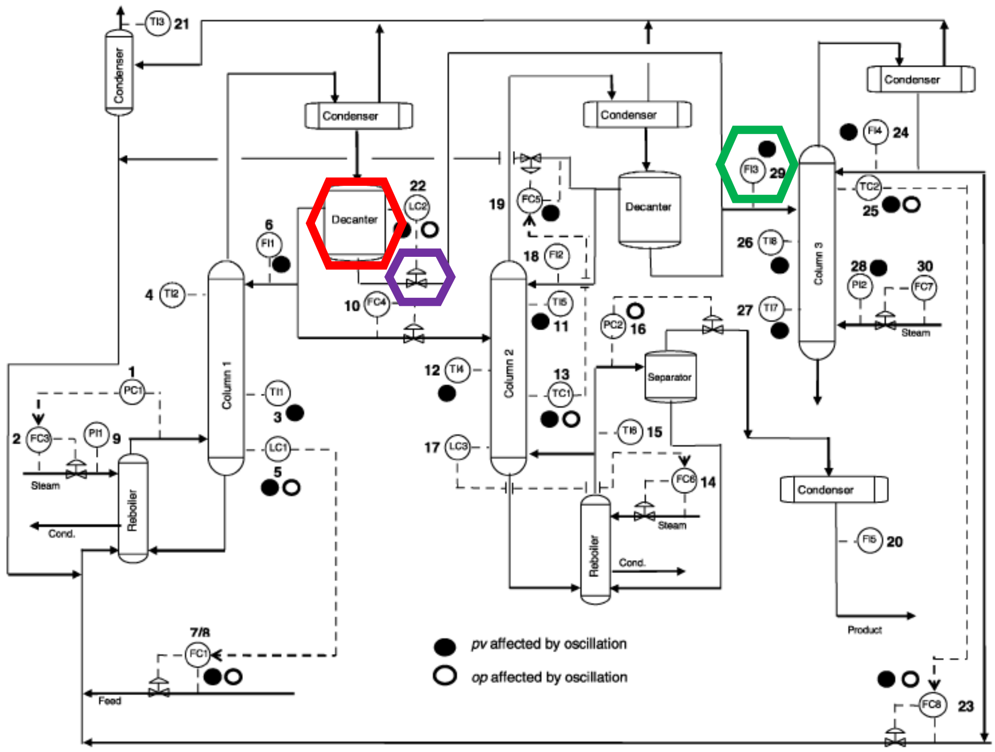


Figure 2: Process schematic showing distributed plant-wide disturbances as a result valve stiction in tag 22 (Thornhill, *et al.*, 2003).

The time trends of the process measurements in Figure 3 show that other tags are oscillating as well. The challenge is to verify that there is a mechanism for the oscillation to propagate plant-wide from the suspected root-cause candidate suggested by data-driven root-cause diagnosis LC2 (tag 22) and affect other process variables as indicated in the data analysis and time trends. The root-cause diagnosis procedure is not complete until a feasible mechanism of oscillation propagation to all the tags suffering from secondary oscillations is explained (Thornhill, *et al.*, 2003). The knowledge about process fluid flow path that eventually led to a decision about proxy measurement was gathered by manual inspection of the process schematic. For large processes, this analysis can quickly become very complicated and challenging to perform manually. Thus an automated approach would be much more suitable and desirable.

The interconnectivity and highly coupled nature of process plants means a localized process disturbance inevitably become a plant-wide problem due to the mechanisms of cause-and-effect and fault propagation along process fluid flow e.g. in recycle streams and control signals lines.

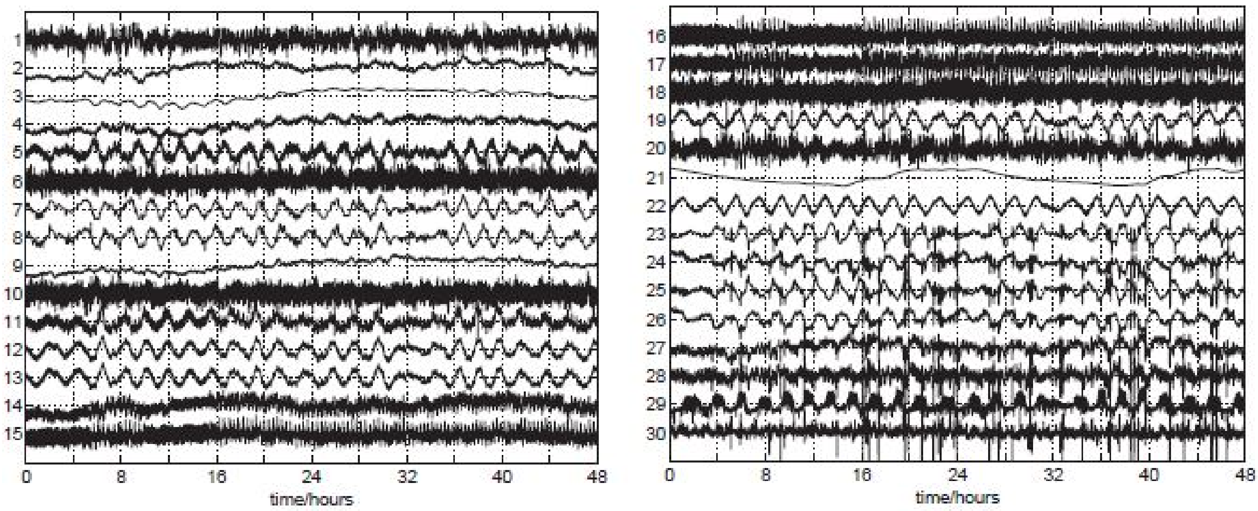


Figure 3: Time trends of process variables of the process described by Thornhill, *et al*., (2003). The plots show a plant with a troublesome plant-wide oscillation.

For example, Figure 4 shows two possible propagation paths from the decanter labelled *'A'*. The path in red colour shows the propagation sequence from the suspected root-cause tag 22.

Uneven flow through the control valve of LC2 would propagate and affect variables along the red path, for example tag 29 and some other variables through recycle streams. The path coloured blue indicates another possible independent propagation path from the decanter.

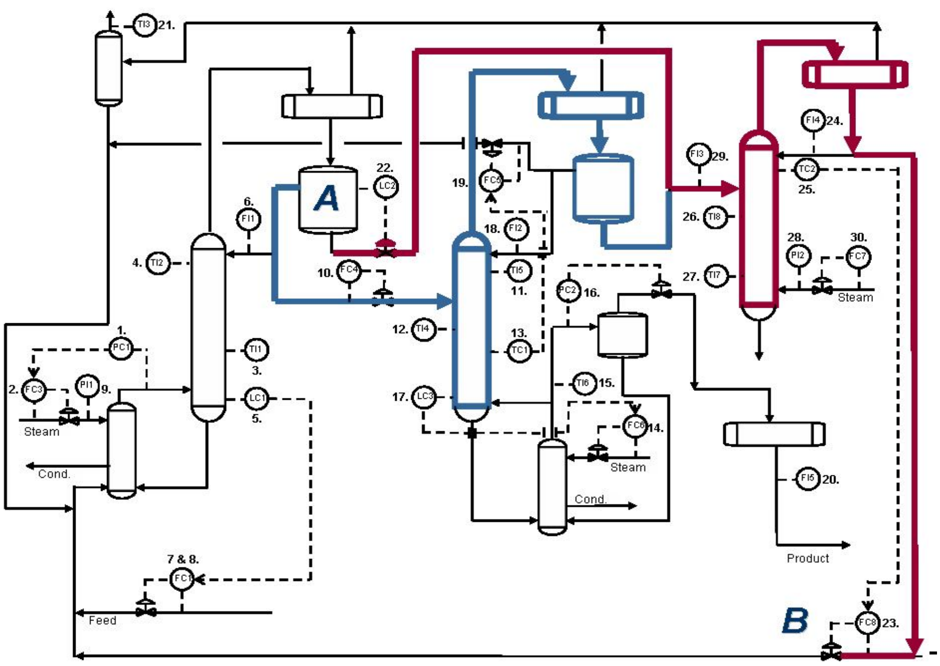


Figure 4: Possible disturbance propagation paths (Yim, *et al*., 2006)

The existence and confirmation of propagation paths among disturbed variables help narrow down some of the hypothesis generated by the data-driven root-cause analysis. The conclusion from the illustrative example above from Thornhill, *et al*., (2003) is that knowledge about process schematic is needed to complement data-driven analysis in order to complete the diagnosis.

1.2 Motivation

One of the key benefits of isolating root-causes of process upsets such as oscillations that manifest plant-wide is that it focuses maintenance efforts on the appropriate process equipment or control loop that needs it and avoids fire fighting or trial and error approaches (Thornhill, *et al*., 2003). This subsection discusses the motivation behind the research project.

1.2.1 Overview

As shown in section 1.1, a purely data-driven technique for plant-wide root-cause diagnosis is incomplete without the information on the overall structural and connectivity information of the process under consideration. In order to complete the diagnosis in the illustrative example above, information about the process structure and connectivity was manually combined with the data-driven result, for example in choosing a proxy measurement point for the fluid flow through LC2 valve.

Traditionally, process control engineers referred to the physical layout of a plant by reading static/printed process schematics, tracing possible causes to effects. However, recently, novel ways of capturing pertinent information about process connectivity with directionality information have been devised and developed (Fedai and Drath, 2005; Laud, 2011). The techniques prescribed and developed allow information about the connections and directions among various process plant items to be captured in an electronic format amenable to manipulation algorithmically in a computer program.

As originally pointed out by (Mohindra and Clark, 1993), an automated means of process diagnosis is appealing because an automatic diagnostic system delivers consistent and reliable performance in the face of complexity. The process models and reasoning methods are explicitly known so that systematic analysis can be used to find and correct malfunctions. Since then, many other authors have commented on the subject including (Kankar, *et al*., 2011; Nandi, *et al*., 2005; Thornhill, *et al*., 2003; Thornhill and Horch, 2007;). Equipping process plant operators and control engineers with tools that will automate and facilitate this manual process of path tracing will save considerable human effort and equipment downtime. As highlighted by Nimmo, (1995), the U.S. economy is losing at least $20B annually from preventable losses from unexpected process disruptions. Automating process diagnosis using connectivity information is at the core of the research reported in this thesis.

The ability of process plant operators and control engineers to make quick and correct decisions in identifying and isolating root-cause(s) of a plant-wide disturbance will save considerable human efforts and equipment downtime. As highlighted by Pinotti, *et al*., (2008); Li, *et al*., (2007) and Cochran, *et al.,* (2011), the time saved translates directly to economic gains. Edwards and Whitaker (2007) reported that more than eighty percent of network downtime is spent looking for root-causes of network problems, while less than twenty percent is spent actually fixing them. The findings highlight the attendant economic implications associated with effective root-cause diagnosis

Connectivity information from a process schematic combined with data-driven techniques has been successfully demonstrated and utilised in diagnostic applications (Scherf, 2006; Yim, *et al*., 2006). Thambirajah (2009) described a parser application that extracts plant items and connectivity information from process schematics using the computer aided engineering exchange (CAEX) standard. The tool was able to find physical propagation paths between two chosen plant elements. This thesis built upon and extended these three previous works using ISO15926 standard and incorporated process know-how on an industrial scale.

The approach is to capture connectivity and directionality information derived from process schematic such as P&ID. Such information will provide useful insights into possible disturbance propagation paths and direction. The connectivity and directionality information combined with results from signal analysis tools prune down spurious statistical correlations among plant variables (hypothesis) suggested by data-driven tools alone.

One of the major challenges is to capture connectivity and directionality information automatically from available sources such as P&ID. There is neither *de facto* nor *de jure* standard for representing process information across the process industry even not among sites within the same organisation. For example, various symbols and standards, such as the ISA-5.1-1984 (R1992) as well as proprietary industry standards are used for P&ID drafting. There is also the problem of dealing with legacy systems.

This thesis describes innovative techniques for enhancing data-driven root-cause diagnosis of disturbances in large chemical processing plants such as a refinery using connectivity and directionality information such as contained in process intelligent piping and instrumentation diagrams (iP&ID). An iP&ID carries extra information about the drawing entities in a database so that such additional information can be extracted and exported. iP&IDs are discussed further in Section 1.2.4 and Section 3.2 of the thesis.

1.2.2 Economic Implication

One way of identifying the potential benefits of detection and elimination of disturbances such as oscillations through improved process control and monitoring strategy is to examine the decreased product variability resulting from the application of process control strategy (Gunther, *et al*., 2007). However, many of the developed and implemented approaches at present such as minimum variance and advanced controller tuning are based on individual control loop and are localised. As pointed out Section 1.1, the propagated effects are often not addressed.

To illustrate how a localised detection and diagnosis of process upset such as oscillation can add benefit to the process, consider Figure 5 (Kinney, 2005) which shows a typical pattern of product variability over time for a process unit. The multiplying effect of the economic implication in a plant-wide disturbance situation will become obvious at the end of this illustration.

As shown in Figure 5, high variability in product quality on the left hand side of the graph prevents the plant from being operated at optimal set point. The limit represents a hard constraint that should not be violated, such as product specification or unacceptable impurity level. Typically, product profitability is inversely proportional to variability in process variables of interest.

Kinney (2005) found process interaction due to product recycle and energy integration responsible for the high variability. The propagated secondary effects can be felt in control loops downstream. Halfway through the graph at about time 1000, the source of plant-wide process interaction was detected and corrective action taken. The variability in product quality is drastically reduced which allows the plant to be operated close to specification limit, thus improving process efficiency. The set point (blue line) is raised closer to the specification limit on the right hand side of the graph with the corresponding decrease in energy usage (green line), thus minimizing production cost and maximizing process profitability.

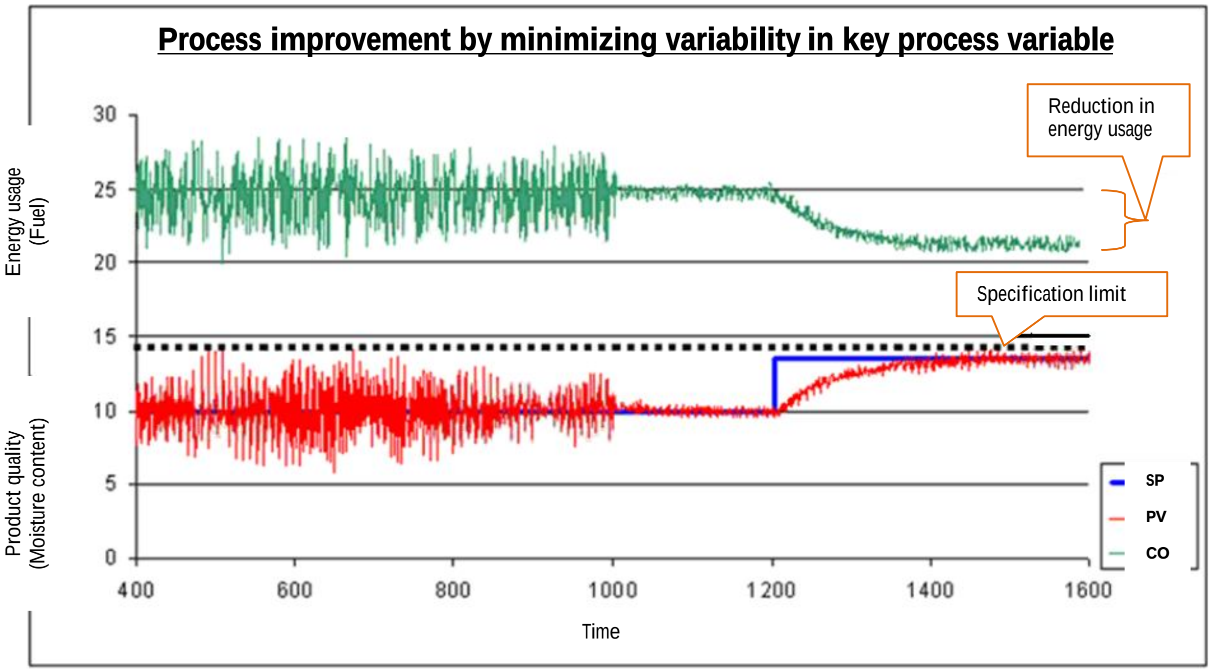


Figure 5: Process optimization by detecting the source of process interaction and taking corrective action. High oscillation on the left hand side results in process inefficiencies. Reduced variability on the right hand side leads to optimal performance, reduced energy usage and thus higher profitability. Adapted from Kinney (2005)

The product quality specification in this drying process case is the moisture content of the product which is set at 14% maximum represented by the dotted line. To avoid violating this limit, the set point (blue line) must be lower than the specification limit so that the typical variation in moisture content does not violate this limit i.e. exceed 14%. With high variability, much of the product is over-dried (moisture content much less than the specified 14%) which ultimately results in high energy consumption. However, with reduced variability on the right hand side of the graph, the plant can be operated at higher set point close enough below the specification limit without violating set constraints. Fuel consumption (fuel flow) is reduced resulting in energy savings and of course, money.

When product quality specification, i.e. moisture content in this example, is less than 14% energy is wasted to produce unwarranted over-dried product while product with moisture content higher that 14% is off specification and must be either recycled or re-blended to meet specification limit. This process of reproducing products of poor quality has production overheads associated with it. The illustrated drying process can be applied to other processes where process variables oscillate. Minimization or elimination of process variability generates significant economic and environmental benefits for the process industries.

This thesis addresses a wider, plant-wide problem as opposed to individual control loops or units by utilizing connectivity and directionality information captured electronically from intelligent process P&ID. The distinction between traditional P&IDs and intelligent P&IDs is highlighted in section 1.2.4.

1.2.3 Requirements for Plant-Wide Diagnosis

Several authors have identified the requirements for a plant-wide approach to control loop performance analysis, detection and diagnosis at an industrial scale (Desborough and Miller, 2002; Paulonis and Cox, 2003; Perry, *et al*., 2000; Qin, 1998). These requirements include:

- Facility-wide benchmarking and standardization of control systems;
- Characterisation of performance faults;
- Detection of the presence of one or more periodic oscillations;
- Detection of non-periodic disturbances and plant upsets;
- Determination of the locations of the various oscillations/disturbances in the plant and their most likely root-causes;
- Incorporation of process knowledge such as the role of each controller;
- Automated model-free causal analysis to find the most likely root-causes.

A number of approaches and techniques in quantitative and data-driven methods have been employed in dealing with the first five of the bullet-point requirements above. For example the methods described in the reviews by Harris, *et al*.(1999), Thornhill and Horch (2007), Venkatasubramanian, *et al*., (2003a), Venkatasubramanian, *et al*., (2003b) and Venkatasubramanian, *et al*., (2003c) provide comprehensive approaches in literature for meeting the industrial requirements for process plant-wide diagnosis listed above. The traditional approach of interpreting results from data-driven analysis by process control engineers can be speeded up and enhanced by automating the process. The automation process ensures logical, consistent and documented reasoning. Requirements 6 and 7 have been demonstrated in prototype tools (Scherf, 2006; Thambirajah, *et al*., 2009; Yim, *et al*., 2006). This thesis describes a much larger, industrial scale plant-wide approach using vendor and platform independent ISO15926 extensible mark-up language (XML) description of process topology.

1.2.4 Traditional "dumb" P&IDs versus "intelligent"/"smart" P&IDs

This section highlights the distinction between traditional, two-dimensional raster-based P&ID graphics and the new generation of intelligent, vector-based, three-dimensional-capable, database-driven P&IDs. The section serves as an introduction to intelligent P&IDs to be discussed in detail later in the thesis in chapter 3.

Intelligent P&IDs (iP&IDs) are relevant to the work of the thesis because they allow electronic capture of process schematics in a platform and vendor independent text-based format that allows for algorithm manipulation by a computer program to generate connectivity matrix automatically. The connectivity matrix shows the directional connections among plant items and utilizes the directional links as contained in the electronic iP&IDs. XML and connectivity matrix are discussed in detail in chapter 3.

Historically, P&ID drawings have contained just two dimensional arcs and lines known as *dumb* P&IDs (Walker, 2009). An example is shown in Figure 6 showing the drawing graphics. Traditional P&IDs are physical sequence of symbols, arcs and lines representing plant's equipment, piping and instrumentation.

There have been significant recent developments in computer aided design (CAD) and drawing methods for P&IDs allowing CAD tools to create P&IDs which are considered intelligent (also referred to as *smart* P&IDs). *Intelligent* P&IDs integrate much more data and information than their *dumb* counterparts.

In addition to the graphic display on the drawing tool, drawings on intelligent P&ID have database connectivity behind them which is a repository of such information as engineering rules, standards compliance, and export of drawings in text-based formats such as XML, automatic design validation, integration with design and calculation packages (DARATECH, 2004). An example of additional information provided by intelligent CAD drawings is shown in Figure 7.

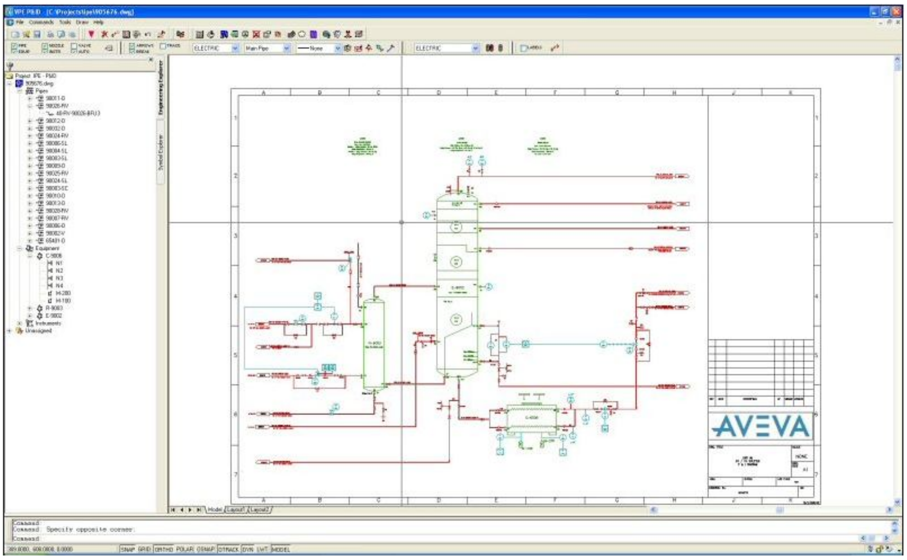


Figure 6: A Piping and Instrumentation Diagram example. (Source http://www.aveva.com/products_services_aveva_plant_pid.php)

Options for converting *dumb* P&IDs to *intelligent* P&IDs are discussed in Section 3.3 of the thesis. Both manual and automated approaches are discussed, hence using intelligent P&IDs in this thesis is not restrictive because most new P&IDs in future will be prepared with intelligent CAD tools, and legacy drawings can be converted. Intelligent P&ID exported as vendor and platform independent XML representation is used as one of the input to *Process Connectivity Analyser* software tool developed in this work.

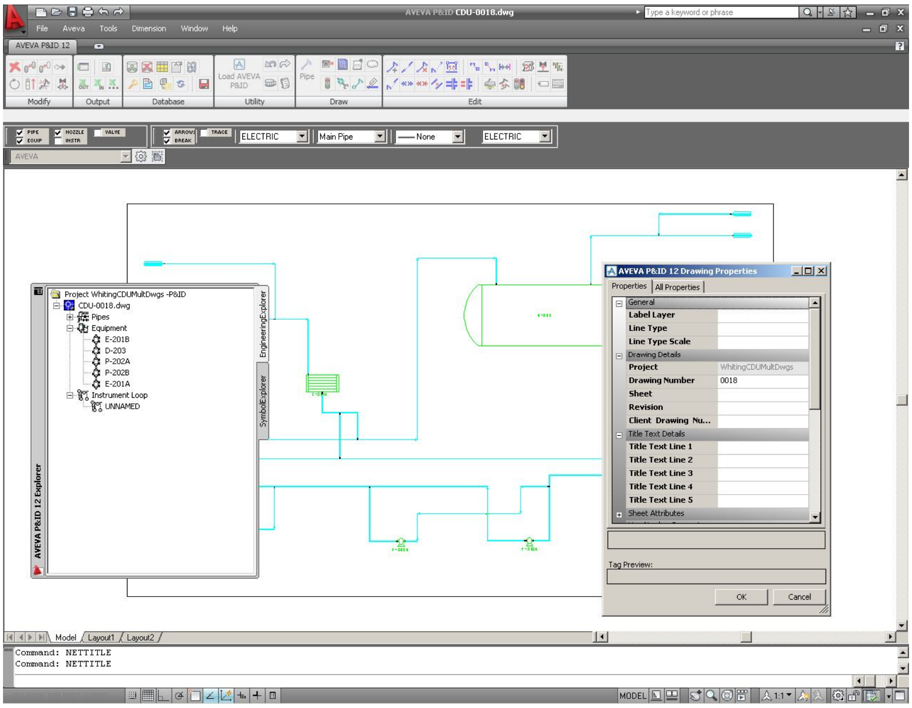


Figure 7: Intelligent P&IDs carry extra information that can be extracted and exported to other software packages. The P&ID was created with AVEVA® P&ID running on AutoCAD®

1.3 Overview of Research Output and Contribution

This section enumerates tasks accomplished in order to meet research objectives and lists specific outputs from the research.

The following steps were taken in order to meet research objectives:

- Survey of the literature in order to understand the scope of the research objectives set out above.
- Comparison of commercial vendors of computer aided design (CAD) tools for creating intelligent P&IDs and conversion of dumb P&IDs to intelligent P&IDs.
- Research into the outputs from various CAD tools to ensure that they conform to international standards such as ISO15926.
- Coding of a parser to construct connectivity matrix from a text-based, vendor and platform independent extensible mark-up language (XML) description of the process. An example of XML is shown in Figure 9.

- Devising of algorithms that link the connectivity matrix with results of cause-and-effect analysis of process data and process know-how and understanding to gain insights into operation of the process and generate logical conclusion.
- Creation of a documented application program in Windows environment (Microsoft .NET) to test and validate research findings (see Figure 12, Figure 13, Figure 14, Figure 16 and Figure 19 for screenshots of the tool developed).

1.3.1 Research Output

A considerable amount of time and efforts was spent on software engineering process because the project sponsor (BP) required a non-trivial software deliverable at the end of the research project.

Consequently, research objectives enumerated in the introductory section of chapter 1 have been analysed and implemented as a software tool. The main components of the software tool and data input sources are shown in integrated block diagram in Figure 8.

The section begins with a description of the input data, namely an XML file and results from pure data-driven analysis. This is followed by a brief introduction to *Process Connectivity Analyser* software tool also referred to as *connectivity tool*, the final output from the project engineered with the research findings.

Detailed description of the steps involved in using the PDA and *Process Connectivity Analyser* tools are explained in chapter 5

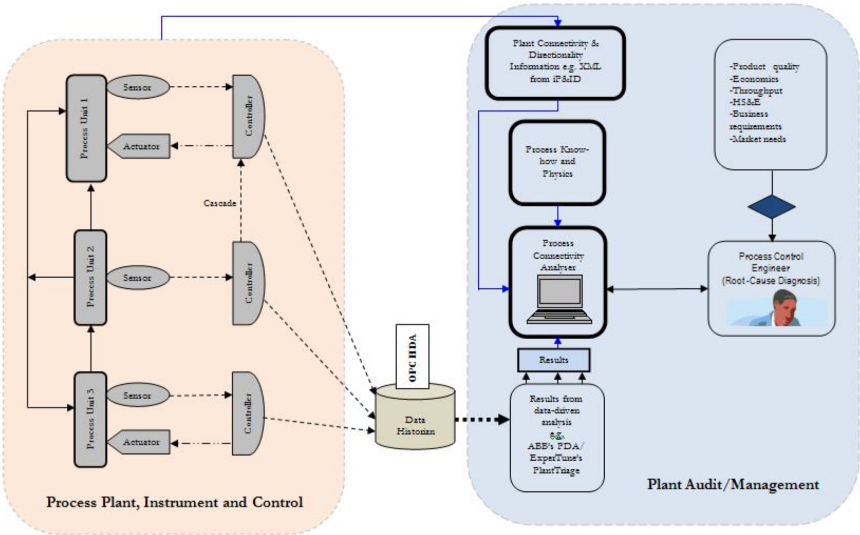
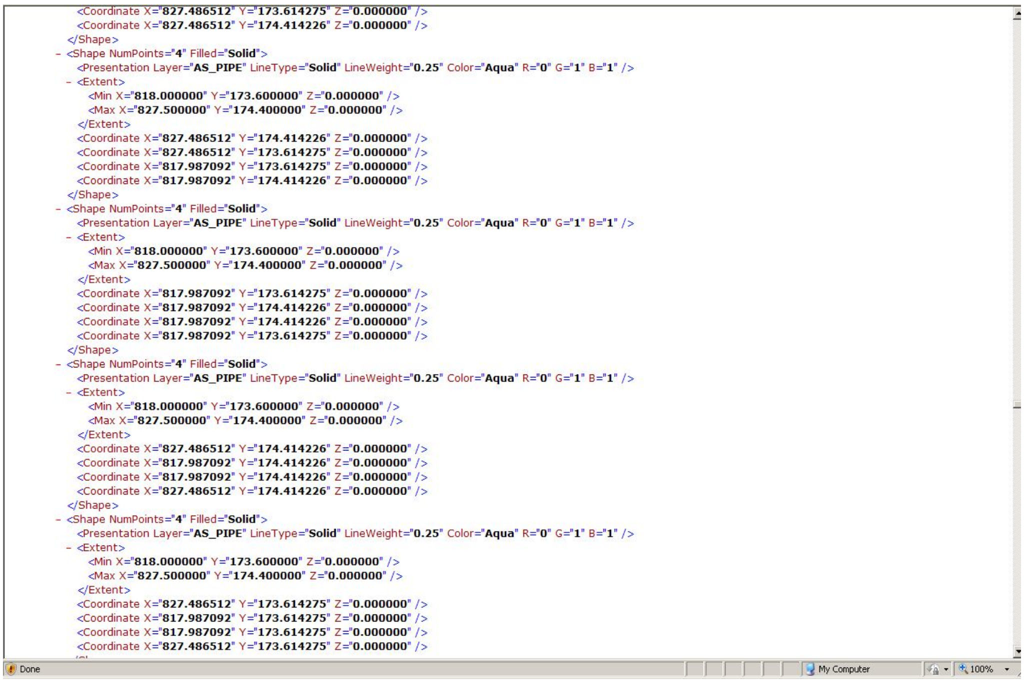


Figure 8: Block diagrams representing the integration of research objectives

XML Input Data File Describing Process Connectivity and Directionality Information from P&ID

An example of process description in XML format is shown in Figure 9. Detailed specification and description of XML is discussed in chapter 3. A standardized, text-based, platform and vendor independent XML file describing plant's items and directional connectivity among them is an input file to the *Process Connectivity Analyser* tool.



```
    <Coordinate X="827.486512" Y="173.614275" Z="0.000000" />
    <Coordinate X="827.486512" Y="174.414226" Z="0.000000" />
  </Shape>
- <Shape NumPoints="4" Filled="Solid">
    <Presentation Layer="AS_PIPE" LineType="Solid" LineWeight="0.25" Color="Aqua" R="0" G="1" B="1" />
  - <Extent>
      <Min X="818.000000" Y="173.600000" Z="0.000000" />
      <Max X="827.500000" Y="174.400000" Z="0.000000" />
    </Extent>
    <Coordinate X="827.486512" Y="174.414226" Z="0.000000" />
    <Coordinate X="827.486512" Y="173.614275" Z="0.000000" />
    <Coordinate X="817.987092" Y="173.614275" Z="0.000000" />
    <Coordinate X="817.987092" Y="174.414226" Z="0.000000" />
  </Shape>
- <Shape NumPoints="4" Filled="Solid">
    <Presentation Layer="AS_PIPE" LineType="Solid" LineWeight="0.25" Color="Aqua" R="0" G="1" B="1" />
  - <Extent>
      <Min X="818.000000" Y="173.600000" Z="0.000000" />
      <Max X="827.500000" Y="174.400000" Z="0.000000" />
    </Extent>
    <Coordinate X="817.987092" Y="173.614275" Z="0.000000" />
    <Coordinate X="817.987092" Y="174.414226" Z="0.000000" />
    <Coordinate X="817.987092" Y="174.414226" Z="0.000000" />
    <Coordinate X="817.987092" Y="173.614275" Z="0.000000" />
  </Shape>
- <Shape NumPoints="4" Filled="Solid">
    <Presentation Layer="AS_PIPE" LineType="Solid" LineWeight="0.25" Color="Aqua" R="0" G="1" B="1" />
  - <Extent>
      <Min X="818.000000" Y="173.600000" Z="0.000000" />
      <Max X="827.500000" Y="174.400000" Z="0.000000" />
    </Extent>
    <Coordinate X="827.486512" Y="174.414226" Z="0.000000" />
    <Coordinate X="817.987092" Y="174.414226" Z="0.000000" />
    <Coordinate X="817.987092" Y="174.414226" Z="0.000000" />
    <Coordinate X="827.486512" Y="174.414226" Z="0.000000" />
  </Shape>
- <Shape NumPoints="4" Filled="Solid">
    <Presentation Layer="AS_PIPE" LineType="Solid" LineWeight="0.25" Color="Aqua" R="0" G="1" B="1" />
  - <Extent>
      <Min X="818.000000" Y="173.600000" Z="0.000000" />
      <Max X="827.500000" Y="174.400000" Z="0.000000" />
    </Extent>
    <Coordinate X="827.486512" Y="173.614275" Z="0.000000" />
    <Coordinate X="817.987092" Y="173.614275" Z="0.000000" />
    <Coordinate X="817.987092" Y="173.614275" Z="0.000000" />
    <Coordinate X="827.486512" Y="173.614275" Z="0.000000" />
```


Figure 9: An example of XML process description

Input Data File Containing Results from Process Historical Data-Driven Analysis

Many signal processing and analysis tools are available in the market to extract useful information from enormous quantities of measurement data generated on regular basis from process plants. Examples of such tools are the PlantTriage® from ExperTune® and those listed in Table 1. Most of these tools acquire real time data from distributed control systems (DCS) of the plant to continuously monitor the state of the plant. The commercial data-driven tools are discussed further in Section 6.1.1 of the thesis.

A number of key performance assessment indices are calculated for each control loop for monitoring several controller properties. Each assessment is calculated at each assessment period and plant operator can select which performance metrics are to be used to create the loop health assessment which gives an indication of the overall health of the control loop and ultimately health of the plant. Process control engineers diagnose malfunctions by examining and comparing the chosen key performance indices (KPIs) with the best achievable standard already in place and requests corrective action on the loops with a poor performance.

Table 1: Commercial software and vendors for process diagnostics and monitoring

Software	Vendor
PlantTriage®	ExperTune®
Loop Performance Manager®	ABB®
LoopScout®	Honeywell®
Control Performance Monitor®	Matrikon® (owned by Honeywell®)
Loop Analysis® (formerly Control Wizard®)	PAS®
PID Watch®	Aspentech®
PDA®	ABB®
Performance Watch®	Invensys®
Control Monitior®	Control Arts®
PCT Loop Optimizer®	ProControl Technology®
PROBEwatch®	ISC®
INTUNE+	ControlSoft
Control Loop Performance	Capstone Technology
Plant ESP	Control Station
DeltaV Insight	Emerson
rCAAM (RoviSys Control Assessment and Monitoring)	RoviSys

The procedures for using KPIs generated by signal processing tools by plant operators and engineers to diagnose malfunction will be automated and combined with connectivity information derived from process representation such as a P&ID to design and develop the proposed root-cause diagnostic tool in the PhD work.

The purpose of historical data analysis is to detect and diagnose oscillations and distributed disturbances across the plant. A wide variety of algorithms such as principal component analysis and commercial data-driven analysis tools are available. The tool for data analysis described and used in this project is a calculation tool for plant disturbance analysis (PDA) based on signal processing algorithms developed at the Imperial/UCL Centre for Process Systems Engineering (Bauer, *et al*., 2007; Bauer and Thornhill, 2008; Thornhill, 2005; Thornhill, *et al*., 2003; Thornhill, *et al*., 2002) and now commercialized by ABB. An example of a PDA screenshot showing results of data analysis is shown in Figure 10.

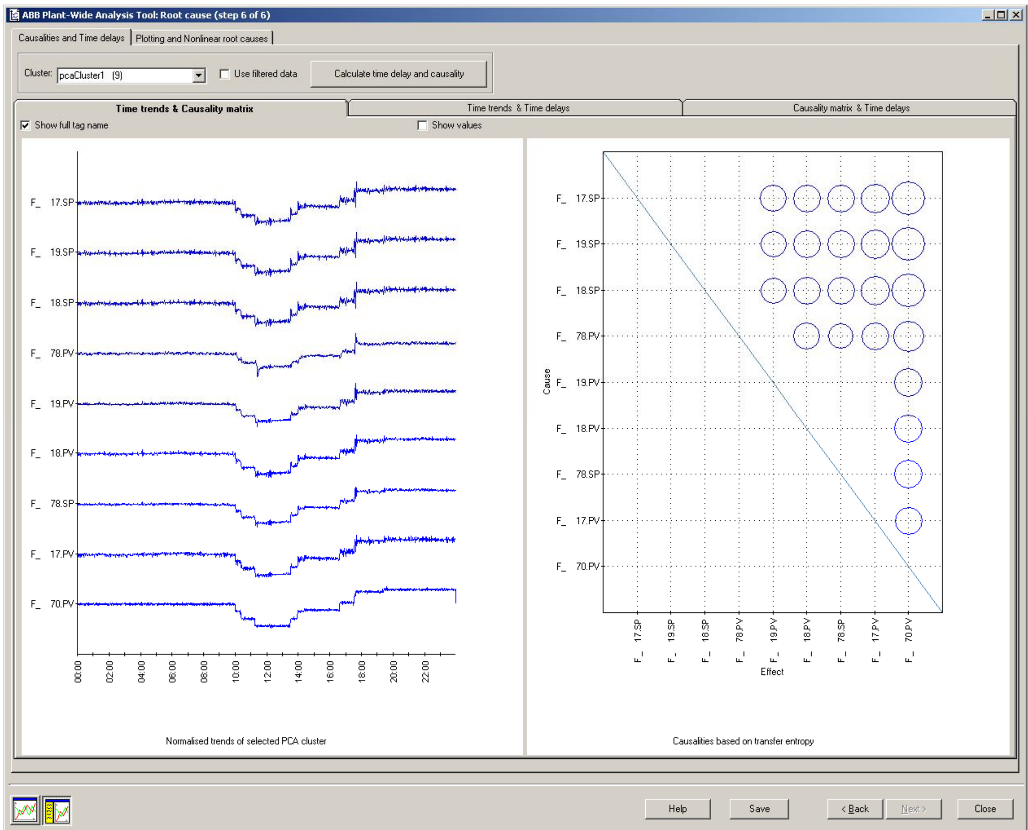


Figure 10: Time series and bubble plot causality analysis suggesting the order of events in the process. This screenshot implies that changes in F_17.SP, F_18.SP, and F_19.SP are the causes of changes in the other tags and that F_70.PV is last in the causal chain.

The reason for using the PDA tools is that PDA is readily available in the university for research. Any other data-driven analysis tool such as PlantTriage® from ExperTune® that seeks to find root-cause of plant-wide disturbance will also suffice. The aim typically is to detect and diagnose root-cause of plant-wide disturbances using purely data-driven analysis. The results produced form the basis for root-cause hypothesis against which test can be carried out to confirm the real root-cause. Spurious results can be eliminated when plant connectivity information from the process schematic is combined with the results from data-driven analysis to enhance the understanding of disturbance propagation through the plant.

The PDA root-cause diagnosis tool is based purely on data because the PDA tool is not capable of knowing the physical layout and relationships between the tags. From experience, the analyses often are *partly* correct, in that the true root-cause is among two or three candidates identified as possibilities. Therefore the diagnosis is generally indicative rather than definitive, and the hypotheses generated by root-cause analysis have to be tested carefully to ensure they make sense. At present, this is done manually by locating the affected tags in the process schematic and applying chemical engineering principles.

The steps involved in data analysis are:

- A spectral cluster analysis is based on the automated comparison of the spectra for detecting similarities, hence it groups tags with similar spectral features. The method used is the spectral principal component analysis as detailed in (Thornhill, *et al*., 2002). Comparing spectra for detecting similarities may be done visually in small scale cases with a small number of tags. In larger scale cases the automated spectral clustering method becomes a necessity.

- A second method looks for clusters of oscillating measurements. The output is a list of clusters of tags characterized by their oscillation period. The oscillation detection uses signal processing methods described in (Thornhill, *et al*., 2003).

Process Connectivity and Graphs

One of the core objectives of the thesis is to demonstrate how connectivity matrix of process plant topology can be generated automatically. A graph shows the various interacting components of a physical system and the connections among them. Graphs have strong theoretical background and have been studied for centuries (Mah, 1983). Graph theory will be discussed in section 2.3.4. A directed graph, a subset of graph, is shown on the right hand side panel in Figure 11, with its connectivity matrix on the left hand panel. The connectivity matrix on the left hand side shows the *nodes* on the rows and columns of the square matrix. The entries in the matrix (*arcs* of the graph) are either a *one* or *zero*, depending on whether there is a direct connection between the elements of the intersection. A *one* indicates direct connection between the element at the row and column of the intersection while a *zero* indicates no connection.

	A	B	C	D	E	F	G	H	I	J	K	L
A	0	0	0	0	0	0	0	0	0	0	0	1
B	0	0	1	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	1	0	0	0	0	0
D	1	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	1	0	0	0	0	0	1	0	0
F	0	1	1	0	0	0	0	1	0	1	0	0
G	0	0	1	0	0	0	0	0	0	0	0	0
H	0	0	1	0	0	0	0	0	0	0	0	1
I	0	0	0	0	0	1	0	0	0	0	0	0
J	0	1	0	0	0	0	0	0	1	0	0	0
K	0	1	0	0	1	0	0	0	0	1	0	0
L	0	0	0	0	0	1	0	0	1	0	0	0

Connectivity matrix

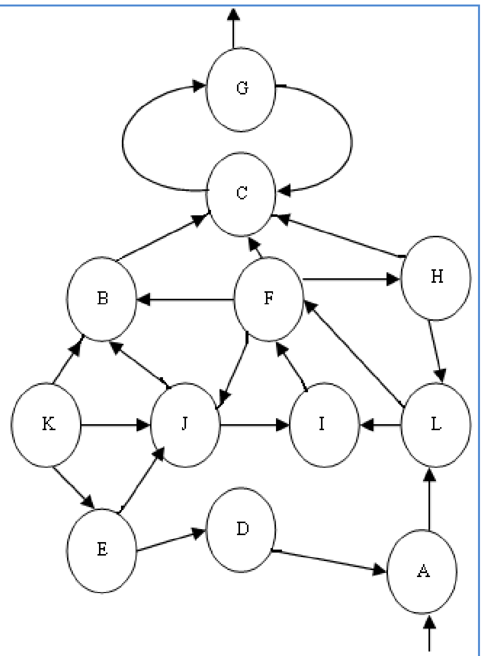


Figure 11: Connectivity matrix created manually from a directed graph on the right hand side

The connectivity matrix is a practical and convenient way to store and manipulate topology of the process plant algorithmically. However, the greatest challenge is how to automatically generate the connectivity matrix of large and complex real life processes. This has been accomplished in this work with the possibility to export connectivity matrix of very large processes to Excel application as shown in Figure 15.

1.3.2 An Introduction to Process Connectivity Analyser Tool

This section provides an overview of the *Process Connectivity Analyser* tool developed from the findings of the thesis. Detail description and operational procedures are presented in Section 5.3. *Process Connectivity Analyser* tool is the main deliverable required by the project sponsor.

Figure 12 shows the start-up/home page of *Process Connectivity Analyser*, the tool developed. A list of basic steps to be followed in using the tool for analysis is presented on the first window. The start up window also displays a calendar with the current date highlighted.

Upon request, the navigation-based *Process Connectivity Analyser* presents the user with the main control panel shown in Figure 13.

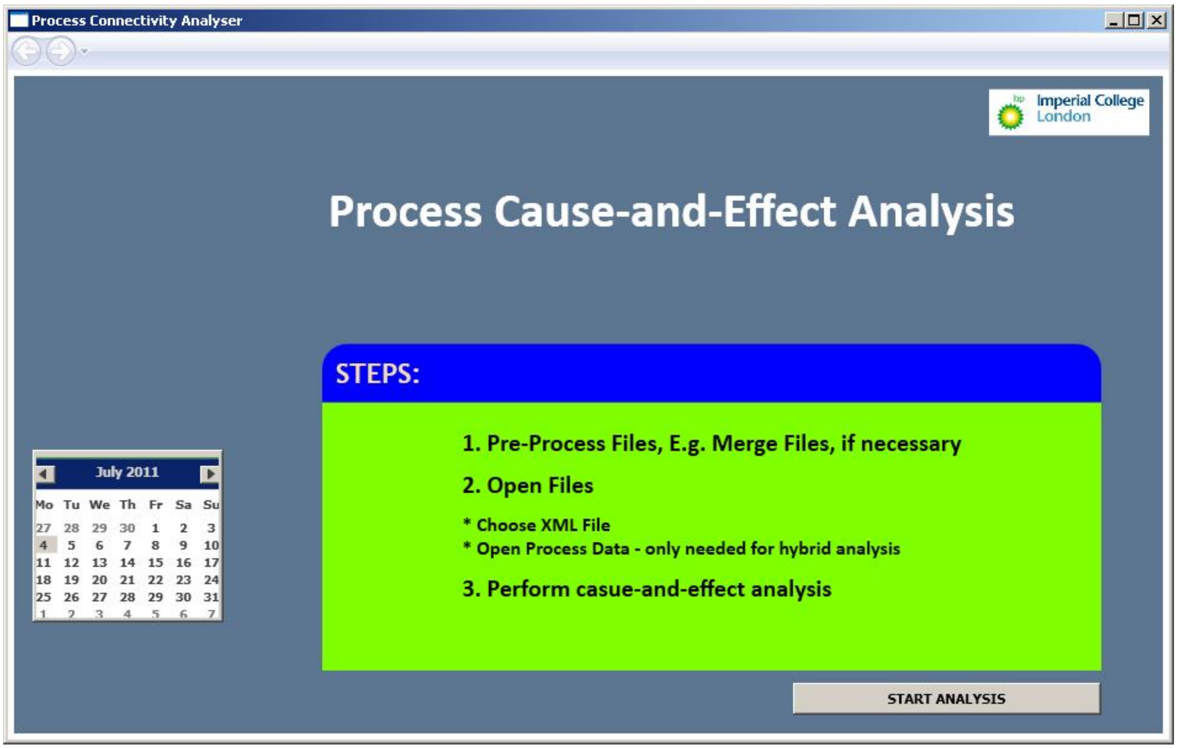


Figure 12: Start-up window of *Process Connectivity Analyser* tool

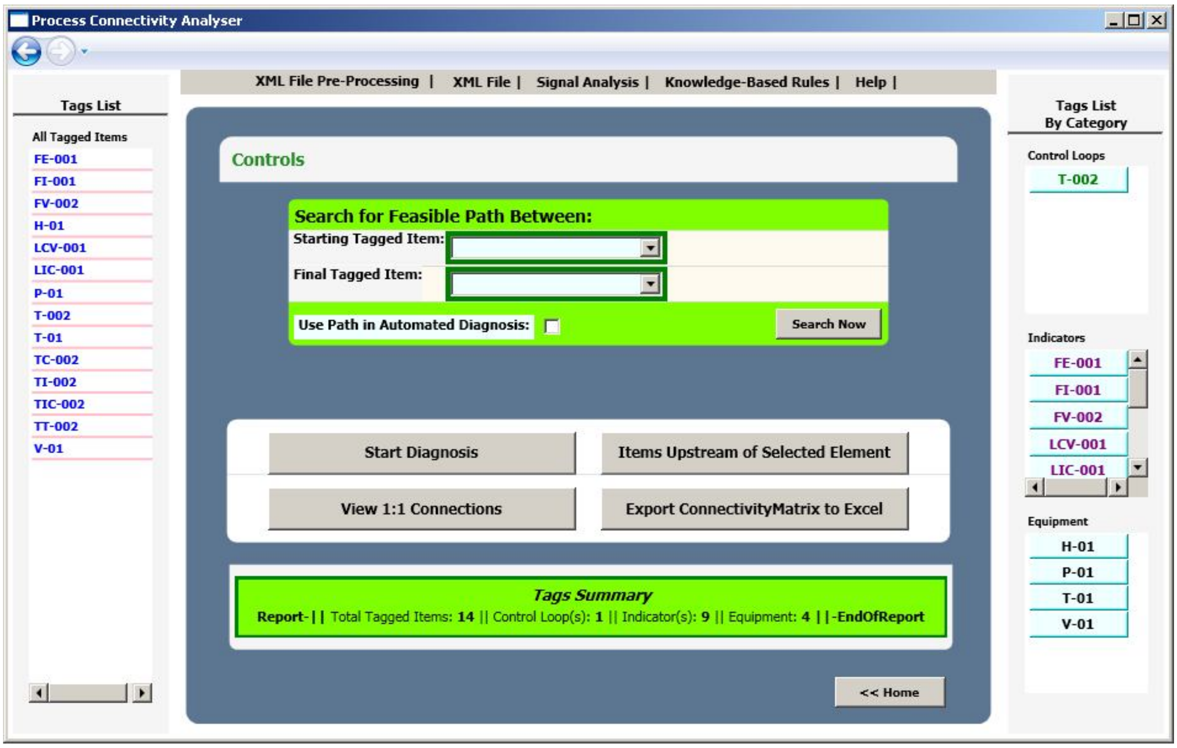


Figure 13: *Process Connectivity Analyser's* main control panel

The user interface shown in Figure 13 allows the user to carry out major operations such as file(s) loading, physical path-finding, root-cause diagnosis, export of connectivity matrix to Excel application and running a check on the original P&ID drawing to ensure that all plant items drawn are fully connected.

The window also displays plant items sorted into categories as equipment, controller or indicator. A search conducted to check physical propagation path between two chosen plant items is displayed on another window as shown in Figure 14.

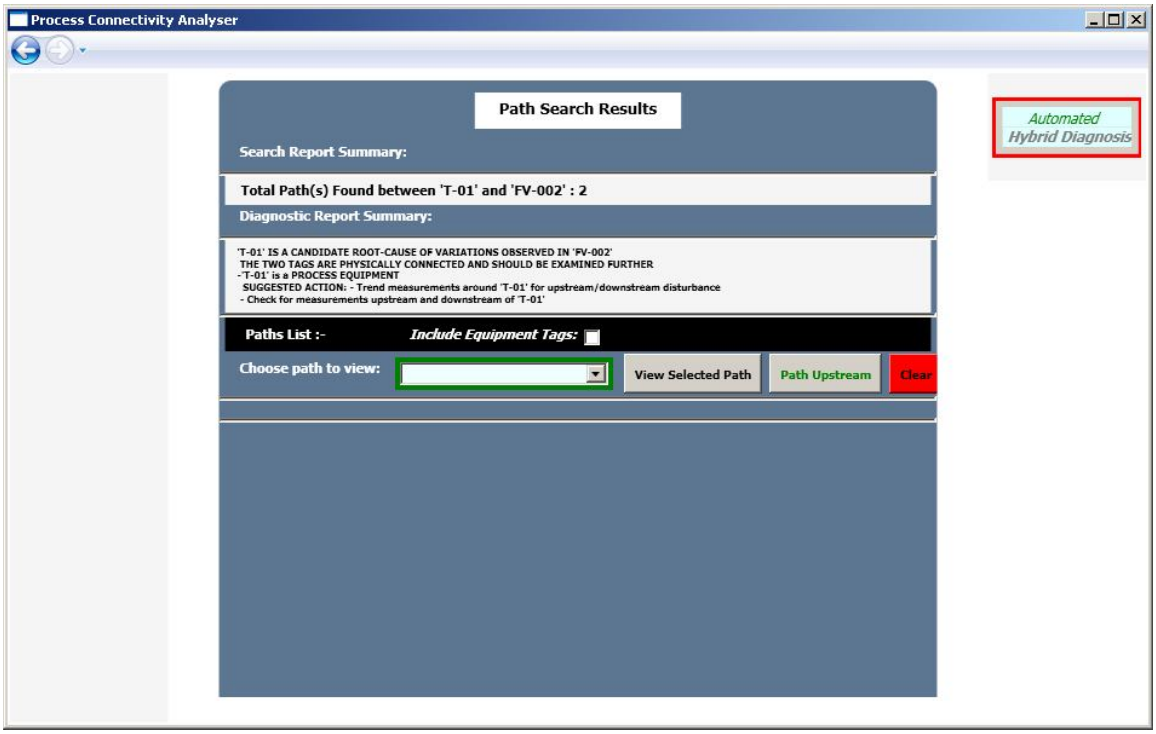
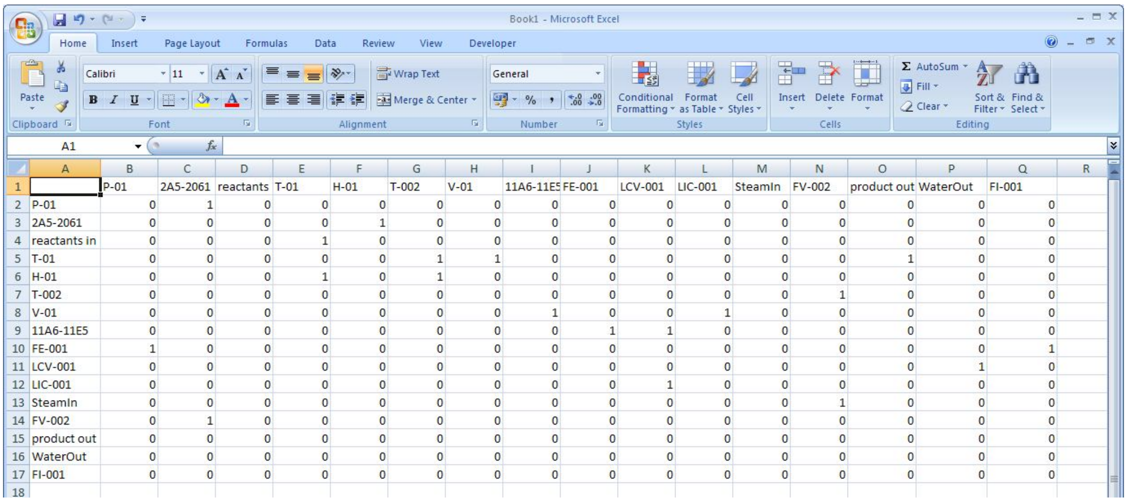


Figure 14: A search result from one plant item to another to check for physical propagation path



	P-01	2A5-2061	reactants	T-01	H-01	T-002	V-01	11A6-11E5	FE-001	LCV-001	LIC-001	SteamIn	FV-002	product out	WaterOut	FI-001
P-01	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2A5-2061	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
reactants in	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
T-01	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0
H-01	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
T-002	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
V-01	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
11A6-11E5	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
FE-001	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
LCV-001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
LIC-001	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
SteamIn	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
FV-002	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
product out	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WaterOut	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FI-001	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 15: Automated connectivity matrix exported by *Process Connectivity Analyser* tool as a stand-alone module to Microsoft® Excel© application

Figure 16 is a separate window in *Process Connectivity Analyser* for conducting diagnostic and other analysis. It allows the user to combine process topology information with statistical-based results from data-driven analysis.

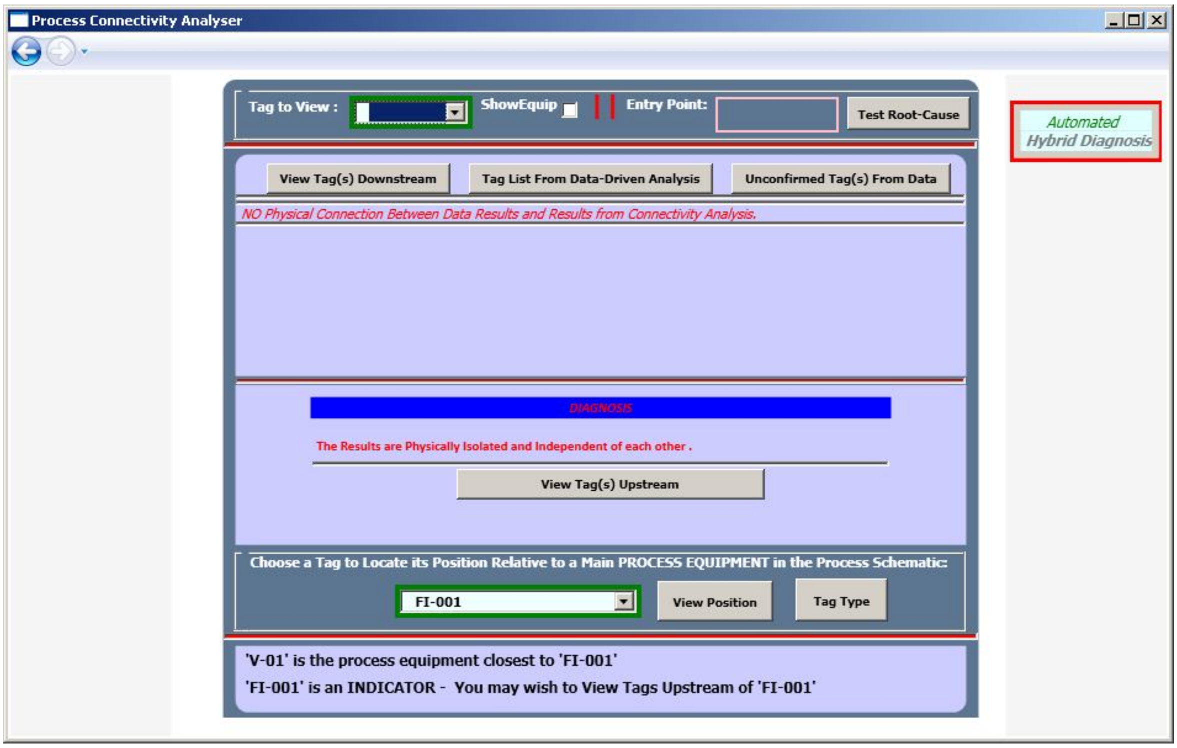


Figure 16: Root-cause hypothesis testing window

Figure 19 provides a facility to cross check process drawing for any loose ends. A loose end is encountered when a drawing entity and piping system appear to be physically connected on visual inspection but there is no actual physical connection upon export as an XML. A simplified and magnified illustration is provided in Figure 17. This might look easy to pick up manually in this illustration but in a real process P&IDs, it is almost impossible to manually detect these gaps.

This can be due to the drawing scale making all connection invisible to human eyes or simply an error in the drawing. Such incorrect drawings with loose connections produce wrong/misleading conclusions in the analysis. It is important to check for this before using the XML export. This facility is provided by *Process Connectivity Analyser* through the user interface shown in Figure 18 and Figure 19.

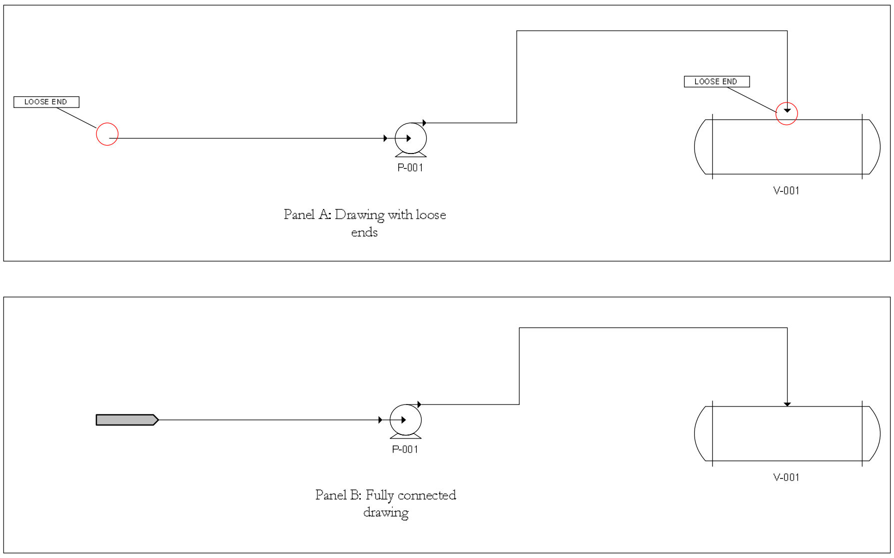


Figure 17: Process schematic drawing illustrating existence of loose ends (top panel) and fully connected drawing (lower panel)

Figure 18 shows a scenario where the connectivity tool detected two loose ends in a drawing and presented the findings as *unidentifiedItem3* and *unidentified4* while Figure 19 depicts a fully connected drawing.

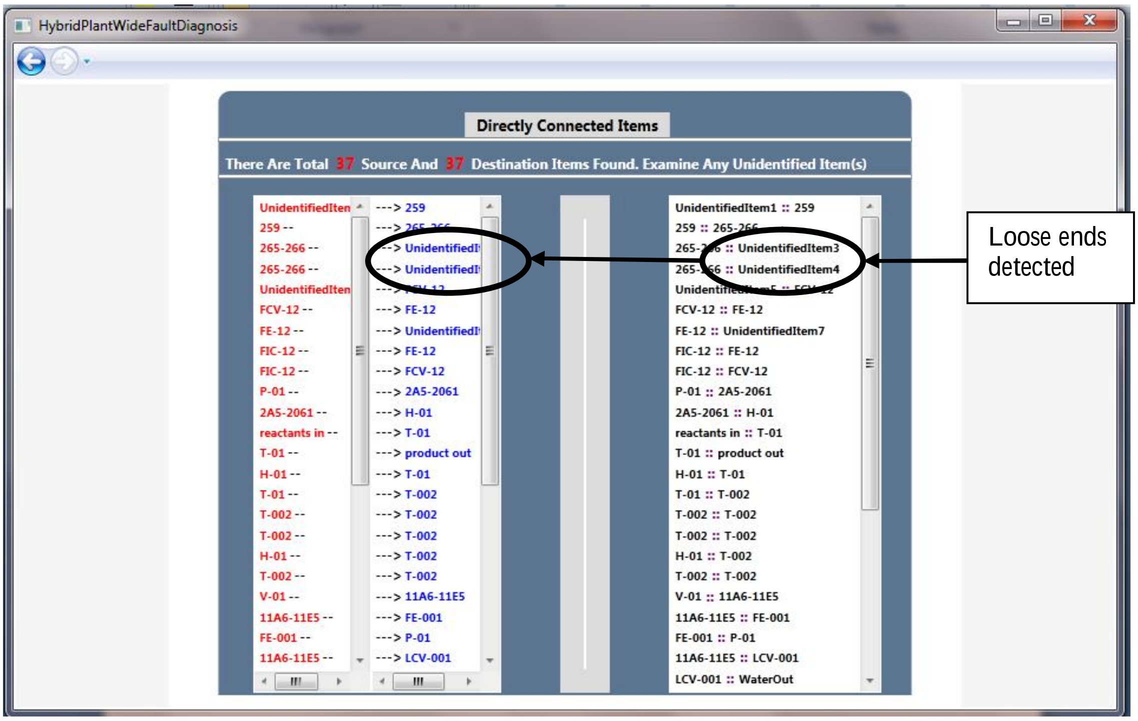


Figure 18: An example of loose end detection due to incomplete drawing of exported P&ID drawing

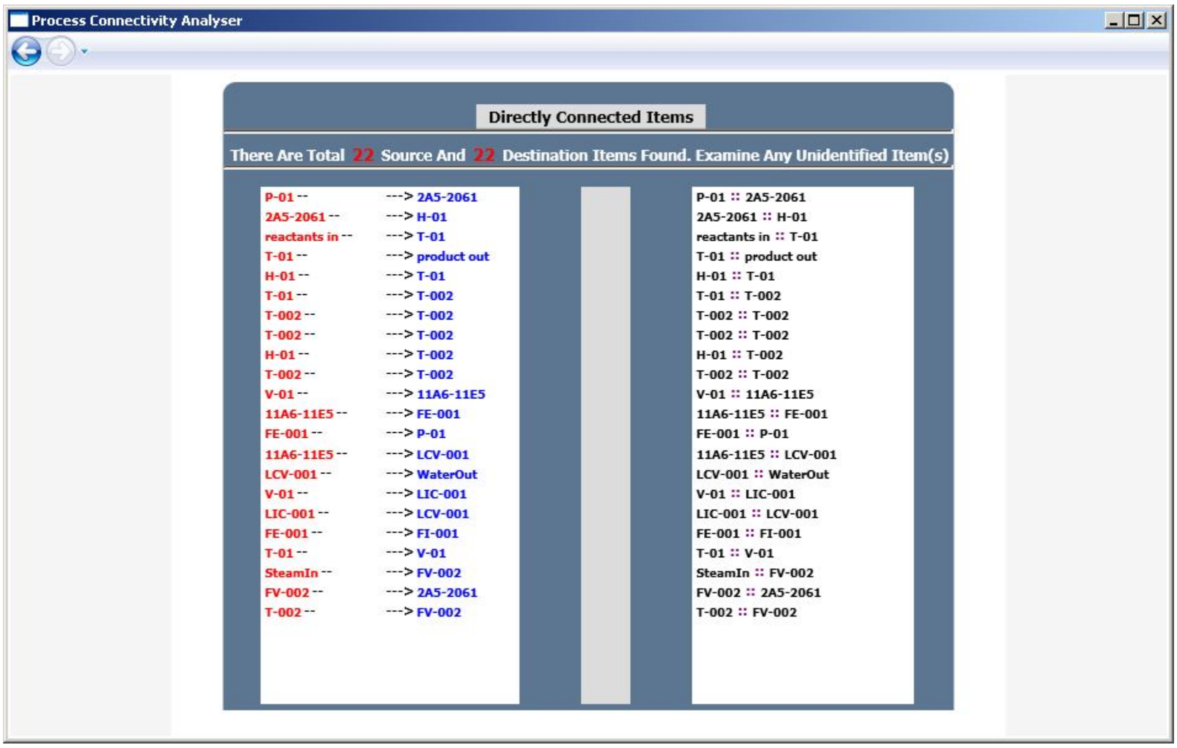


Figure 19: Facility for checking exported P&ID drawing to ensure all plant items are connected with appropriate directionality information

1.3.3 Research Scope

The primary objective of the thesis is to establish new and innovative ways to capture and algorithmically manipulate large electronic process plant topology and interconnections among plant items such as a crude oil refinery for the purpose of diagnosing root-cause of plant-wide disturbances. Process topology is commonly contained in engineering design and construction document such as a P&ID.

Continuous processes are designed to be operated around the steady state point such that any deviation from the steady state but within some allowable limits (normal) can be handled by implemented control systems. In some cases, corrective actions taken by one control system simply upsets another process variable, triggering corrective action from another controller. The overall effect is a persistent dynamic cycle. Controller output at limit and valve stiction or hysteresis will lead to similar sustained disturbance that could propagate plant-wide. In the case of shut-down, start-up or abnormal operations, the deviations from the normal operating point are considered not to be small hence, it is necessary to install devices to handle such situations in addition to applying intuition, experience and sound engineering judgement. Such devices include blowdown systems, start-up heaters or hazard prevention systems(Umeda, *et al*., 1980). The thesis finds application in persistent, dynamic normal plant-wide upset.

The process models generated from process P&IDs are qualitative and static in nature. To take its full advantage however, a model that encapsulates dynamic behaviour of the plant has to be incorporated such as signal-based empirical analysis as well as knowledge about the process as demonstrated by (Di Geronimo Gil, 2010; Thambirajah, *et al*., 2009; Yim, *et al*., 2006).

In addition to the research's primary objective of utilizing topology and connectivity information for root-cause diagnosis, the output from the research would find alternative uses as enumerated in Section 7.3 of the thesis. For example, the tool developed from the research can be used to find which part of the production plant that would be affected when a unit such as a feed pump was shut off prior to carrying out a maintenance or repair work on the plant.

1.3.4 Research Contributions

Key contributions of the research are enumerated below.

- A fully documented software tool in a Windows environment. This is the key requirement from the project sponsor.
- An automated, practical, large-scale, plant-wide process plant topology (ISO15926 based intelligent P&ID) and connectivity information extraction and representation in the form of connectivity matrix. The connectivity matrix is amenable to computational manipulation and could serve as precursor for other applications. For example in automated calculation of degree of freedom analysis(Alabi, 2010) for control structure design and evaluation.
- An automated electronic processing of P&ID for cause-and-effect analysis obviates the need for manual examination of printed P&IDs to get process insight and locate root-causes of process upset. Traditional signal analysis methods only consider measurements points (indicators). The conventional approach will require paper

tracing on process P&ID to locate process equipment or controllers closest to the observed disturbed measurement points as the possible source of disturbance.

- A modular approach allows for as many process P&IDs as necessary to be merged for analysis, creating a true plant-wide analysis for large-scale plants.
- Research findings opening up opportunities to integrate process structural representation with traditional process models for deeper process insights (Di Geronimo Gil, 2010).
- Connectivity information capturing process variables (measurement points) as well as processing equipment. Past researched approaches considered cause-and-effect relationships among process variables. Faults, however, typically don't emanate from passive plant elements such as a measuring instrument but from processing units necessitating manual examination of process P&ID for equipment closest to the measured root-cause upset.
- Demonstration of the possibilities of mixing process qualitative models with other process models.
- Showing that research findings can be engineered into a user-friendly software tool to demonstrate research key concepts and algorithm.

1.3.5 Conference and Journal Papers

In addition to writing three technical reports, several presentations slides and numerous meeting minutes for the industrial sponsor, the research has also produced/in the process of producing the following conference and journal paper:

- Di Geronimo Gil, G.J., Alabi, D.B., Iyun, O.E. and Thornhill, N.F., 2011, Merging process models and plant topology, *Advanced Control of Industrial Processes* (*ADCONIP* 2011), Hangzhou, China, May 23-26 2011.
- A journal paper on the Tennessee Eastman Process (in progress).

1.4 Outline of the Thesis

This chapter has provided a broad overview of the research problem, the motivation and challenges. The chapter also introduced the concept of *dumb* and *intelligent* P&IDs and described high level features of the software tool developed from the research. The chapter concluded with the key contributions of the research work.

- **Chapter two** reviews existing literature on process modelling approaches and application in process plant diagnosis. The chapter considers model-based approaches-quantitative and qualitative as well as analysis of process signal historical data.
  The chapter describes the concept of graph theory, its application, connectivity matrix, reachability matrix, intelligent piping and instrumentation diagrams and XML. It considers techniques from artificial intelligence methods relevant to the thesis. The chapter also discusses the concept of model mixing (hybrid systems) as a way of compensating for shortcomings in individual modelling approach. A table presenting world experts and centres of excellence of the main players in fault detection and diagnosis of chemical plants and the main topics of current interest is presented.

- **Chapter three** focuses on extraction of engineering information from process P&IDs for operational purposes. It also enumerates various industry standards and commercial vendors of intelligent P&IDs. Analytical and iterative steps followed in choosing an intelligent CAD tool for generating ISO15926 compliant XML description of process plants from a variety of commercial vendors based on the needs and requirements of the project are discussed. The chapter presents an illustrative example of a refinery crude heating unit to drive home the points discussed in the chapter and proceeding chapters.

- **Chapter four** describes the software development process. It considers software development lifecycle including requirements definition, requirements specification, design, coding, unit and system testing.

- **Chapter five** begins with an in-depth description of the various functionalities of the software developed from the research findings and a guide on how to use the software. The chapter presents practical demonstrations of the software developed in the project to drive home key concepts and applicability in real life situation using an illustrative example, industrial and academic case studies and discusses major findings.

- **Chapter six** discusses industrial linkages and considers practical application of the tool in real world environment. The chapter considers the potential for the software integration with other software packages-such as those that are purely data-driven. The chapter concludes with alternative uses to which the software developed in the research project can be applied.

- **Chapter seven** distils major conclusions from the report with a summary and suggestions for improvement on the current work and possible future research ideas. The chapter also discusses alternative uses to which the tool developed can be put to within the process industry. There are references in alphabetical order.

1.5 Chapter Summary

This chapter served as an introduction to the topic and layout of the thesis. It highlighted the need and motivation for the research project from both academic and industrial points of view. The chapter also reviewed the various challenges and approaches to be adopted in arriving at solutions to the research problem. A roadmap for the rest of the thesis was provided to aid navigation through the thesis.

2 Literature Review

This chapter presents a review of prior work in order to establish the background for the project from the literature. It explores various approaches used in previous work and the motivation. The relevance of the reviewed papers to the research was to evaluate the various techniques that have been used by various researchers in the field of process analysis and diagnosis in order to gain an insight into and utilize some of the established techniques.

The chapter begins with a high level overview of different techniques of representing physical systems such as a chemical plant in order to carry out analysis, for example for detection and diagnosis of process anomalies, on such systems. Two basic approaches of representing physical systems are considered. These approaches for representing physical system are:

- classifications based the type of output presented, and
- classifications based on the nature of process knowledge available such as the laws of physics.

These two paradigms of process description and analysis are discussed in Section 2.2.
Generally, the distinction in the classifications are based mainly on the source of data used in the analysis, for example those methods utilizing measurements from the process plant and those techniques employing physical structure and process units interconnections are classified differently. A list of various sources of data and information for process description and analysis are provided Section 2.1.1.

Depending on the approach considered, real physical processes can be represented as one or a combination of physical (mock-up), quantitative, qualitative, historical data or mental description for analysis.

A list of experts and centres of excellence pertinent to the research field accompanies the review.

2.1 Process Abstraction and Representation

This section introduces various approaches for representing physical systems. Methods for representing physical systems fall broadly under those that are model-based and model-free. A model is an abstraction or incomplete representation of the real world in order to make the physical system amenable to analysis. For example, a perfect sphere can be used to model the planet earth. The essence of a model is the question or set of questions that the model can reliably answer for us (Buede, 2009).

The section begins by introducing the modelling process and sources of data/information for modelling in chemical process industry. The model and model-free methods of representation are further broken down in a hierarchical order to explain various techniques under the broad categories. The concept of model cross breeding is also discussed.

2.1.1 Process Modelling

In order to study steady state and dynamic behaviour of physical systems it is often convenient to reduce the system into a representation simpler than the original system in order to make the system amenable to analysis. This process of simplification and abstraction of necessary details from real world systems is called modelling. In other words, every model distorts the system under study in order to simplify it. According to Lee (1999), the following three defining characteristics must be present in a model:

- **Representation:** a model needs to represent all important aspects of an underlying system for a given intended purpose.
- **Prediction:** a model needs to allow for estimation of how a system will perform in a given situation.
- **Explanation:** in order to increase confidence in the predicted results of a model, it must be possible to explain or justify their derivation and relate prediction results to the real world.

In most physical systems such as a chemical reaction model, it may be necessary to simplify the process description as long as sufficient details are abstracted from the real system with caution. However, care is required while interpreting the model's predictions to ensure that the model makes sense in describing the real physical system (Puccia and Levins, 1985). To model a system is to replace it by something which is simpler and easier to study and equivalent to the original system in all important respects. Frank at al.(2000) pointed out that models needed for fault detection and isolation can be simpler than those for control as only the pertinent parts of the model which reveal the faults of interest are considered.

In order to establish a model that adequately describes the process under consideration, a wide variety of rich information sources are available to model builders. These include:

- Principles of physical and chemical engineering science
- Piping and Instrumentation Diagram (P&ID)
- Process Flow Diagram (PFD)
- Process flow sheet
- Process history and real time data
- Equipment and instrument specification sheets
- Empirical relationship from regression of data
- Experience of operating personnel
- Event trees and fault trees
- Equipment and instrument specification sheets

The aim in this thesis is to utilize information obtainable from one or a combination of sources from the list above to describe the physical system under consideration.

One of the factors to be considered in choosing one or more data/information sources is the ease with which the data/information can be obtained. Considering the list of sources above, it appears that information contained in process P&ID and historical process measurement data are readily available for carrying out analysis.

Effective practical diagnostic techniques in an industrial setting require a combination of information from various domains such as those listed above. Evidence for this assertion is based on the fact that industrial process P&IDs contain information for analysis and anomaly diagnosis purposes, they are readily available and non-invasive. Export and import of electronic P&IDs are now possible on a wide variety of application such as the AutoCAD®

from Autodesk®. Similarly, real time and offline process measurements data are industrially available and are also non-invasive during analysis. Commercial software tools, such as PlantTriage® from ExperTune® and algorithms, such as principal component analysis (PCA), used for process characterization and dimensionality reduction in huge measurement data space are available and already at advanced levels(Bauer and Thornhill, 2008; Thornhill and Horch, 2007; Thornhill, *et al*., 2002).

Process control engineers and process operators possess a great deal of knowledge about the process system under consideration. This can be further utilized for diagnostic purposes. The underlying physical and chemical laws relating to the process can also be integrated in building an effective diagnostic tool.

2.2 Classification of Process Models

This section discusses and classifies models used in process description and representation based on the:

- final outputs or answers from the model
- type of available knowledge about the process.

2.2.1 Classification Based on Outputs Presented from Models

One way of classifying models is to consider the kind of answers or final presentations such models provide (Buede, 2009). Using this criterion, the following categories, depicted in Figure 20, can be identified:

- **Physical models:** present the real world as an entity in three-dimensional space and can be divided into full-scale mock-up, sub-scale mock-up, breadboard and electronic mock- up. Full-scale mock ups are usually used to match the interfaces between systems and components of the system. An example is a chemical pilot plant. Subscale models are commonly used to examine the behaviour of specific issue such as flow in pipes and other chemical process unit operations. A breadboard is a board on which electronic and mechanical prototypes are built and tested.

- **Quantitative models:** provide answers that are numerical. Models can be either analytic, simulation or judgemental in nature. Analytic model is based on underlying systems of equations that can be solved to produce a set of solutions. These solutions can be developed in a closed form. Simulation methods are used to find a numeric solution when analytic methods are not realistic. Judgemental models provide representations of real world outcomes based on expert opinions. Judgemental models could serve as precursor and basis for other quantitative activities.

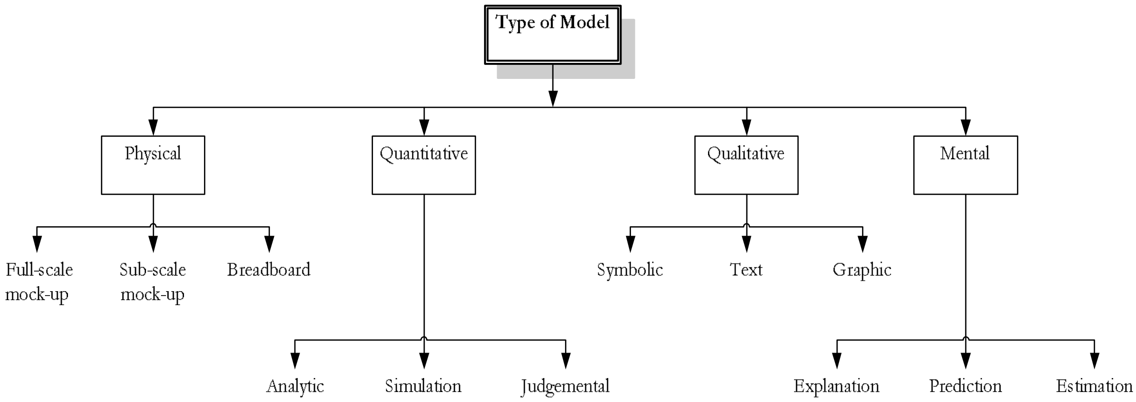


Figure 20: Models classification based on answers they provide

- **Qualitative models:** outputs from qualitative models could be symbolic, textual or graphic. Symbolic models are typically based on logic or set theory. Textual models are based on verbal descriptions of the real world. An example is the use of one or more paragraphs to describe a system's requirements. Graphical models use either elements of mathematical graph theory or simply artistic graphics to represent hierarchical structures or the dynamic interaction of the system's components. The use of artistic graphics as modelling approach is often termed "view graph" engineering. Most engineers consider graphical models as one step above textual models. If graphical models can be based on mathematical graph theory, then these qualitative models can be powerful additions to the analysis toolkit of the system as utilized in this thesis.

- **Mental models:** these are human abstractions of thought such as used by a plant operator for controlling and predicting plants performance and behaviour. Outputs from mental models may lack objectivity as two people may have different mental model of the same real physical system. However, other models, such as quantitative models, are developed through mental process of one or more people and are the product of their mental models. It can be a mistake to ascribe objectivity to models, for example even complex mathematical models often have subjective assumptions throughout their equation and data.

Classification of Models Based on Available Knowledge

The three series reviews by (Venkatasubramanian, *et al*., 2003a; Venkatasubramanian, *et al*., 2003b; Venkatsubramanian, *et al*., 2003c) provide a detailed and comprehensive review of process analysis techniques, with a bias on fault detection and diagnosis, based on available *a priori* knowledge for model-based quantitative and qualitative analysis and *a posteriori* knowledge for data-driven, process history-based methods. The overall hierarchical classification with additions from (Thornhill and Horch, 2007) is shown schematically in Figure 21. The classification approach groups process analysis broadly under model-based and model-free or data-based. Subsequent subsections review the pertinent research work from this classification.

2.3 Model-Based Approach

Model-based approach to physical process representation, such as chemical reaction kinetics, utilizes fundamental knowledge about the system under consideration. Such model development includes:

- First principles approach, which captures process dynamics and are typically based on fundamental underlying physics of the process being analysed.
- State-space model or Transfer function, which is a linear representation of the input-output relationship of the underlying physical system under consideration.

Model-based approach to process description can be classified under two broad categories as quantitative and qualitative. This subsection reviews the various model-based process modelling paradigms with emphasis on fault detection and diagnosis application.

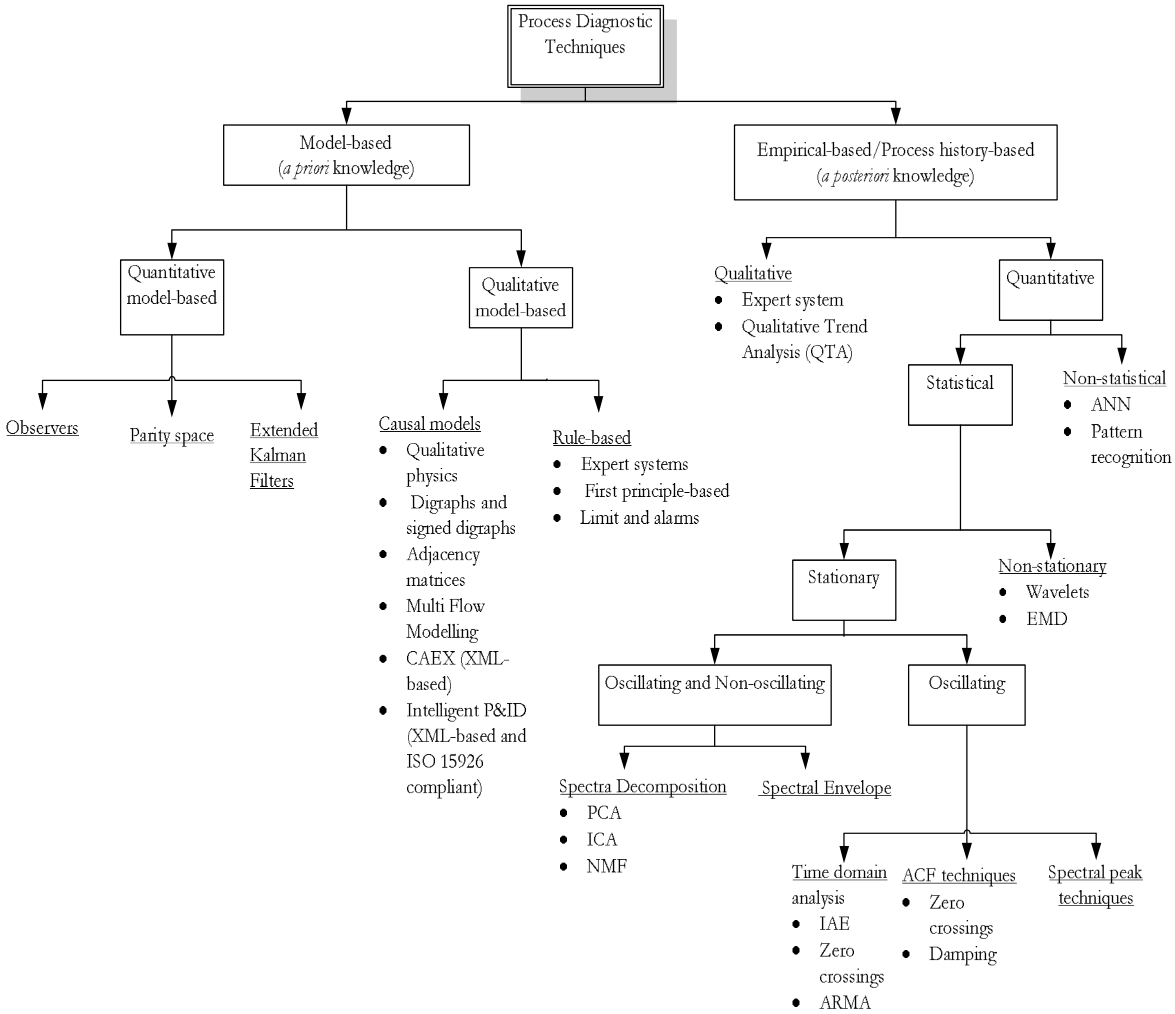


Figure 21: Overall hierarchical classification of process modelling and analysis. (Thornhill and Horch, 2007; Venkatasubramanian, *et al*., 2003a; Venkatasubramanian, *et al*., 2003b; Venkatsubramanian, *et al*., 2003c)

2.3.1 Quantitative Model-Based Methods

Quantitative model-based methods have been traditionally used to model physical and engineering systems. Depending on the purpose of the model, physical models derived quantitatively could be detailed or simplified. Building quantitative models of physical and engineering systems require *a priori* fundamental understanding of the physical system under consideration. By comparing process measurements with analytically computed values from the process models, any statistically significant deviation that signifies the occurrence of a fault can be identified (Kosebalaban and Cinar, 2001). Some applications of quantitative model-based diagnosis are depicted in Figure 22. Detailed knowledge of the physical relationships and characteristics of all components in a system are represented as a set of mathematical equations based on mass, energy, momentum and stoichiometric balance to build a detailed model of the system under consideration. The model is then used to estimate process parameters.

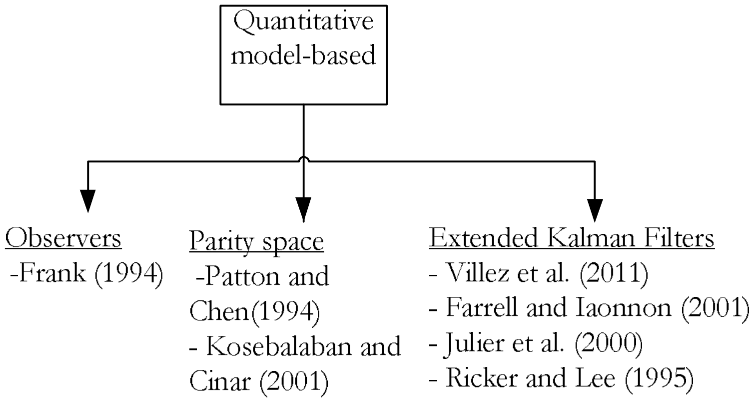


Figure 22: Quantitative model-based techniques for fault detection and diagnosis

Figure 23 depicts one way of classifying quantitative models as a collection of mathematical relationships among process variables. The sets of mathematical equations representing the system are then arranged, solved and interpreted as shown in Figure 24, which summarises the iterative steps involved in developing and using a mathematical process model and simulation.

The power of this method lies in the fact that models are built on sound physical and engineering principles and transient dynamic behaviour of the system can be analysed. Quantitative models also have higher predictive power than qualitative methods.

Over the past two decades, a large number of techniques for plant-wide fault detection have been developed using quantitative or data-driven methods. Prominent among these are:

Observers and Parity Space:

The use of models for detection and diagnosis typically involves comparison of process measurements with values calculated by the model. Parity relations involve checking the values calculated by the process models with the output measurement data from the sensors for any inconsistency or residuals. Theoretically, a smoothly running plant without any disturbance will generate zero residual. However, in real life situations, noise in sensor measurements seldom produces zero residual even in fault-free situation (Kosebalaban and Cinar, 2001).

Hence, it is imperative to apply statistical tests to differentiate deviations due to noise from actual faults. Patton and Chen (1994) provide the state of the art review of parity space fault diagnosis with application in aerospace domain. The paper addresses the issue of robustness in deploying parity space for residual generation across a wide range of application. Frank (1994) reviewed the structural equivalence between parity-based methods and observer-based methods for residual generation and extended the discussion to incorporate enhanced robustness by using adaptive thresholds.

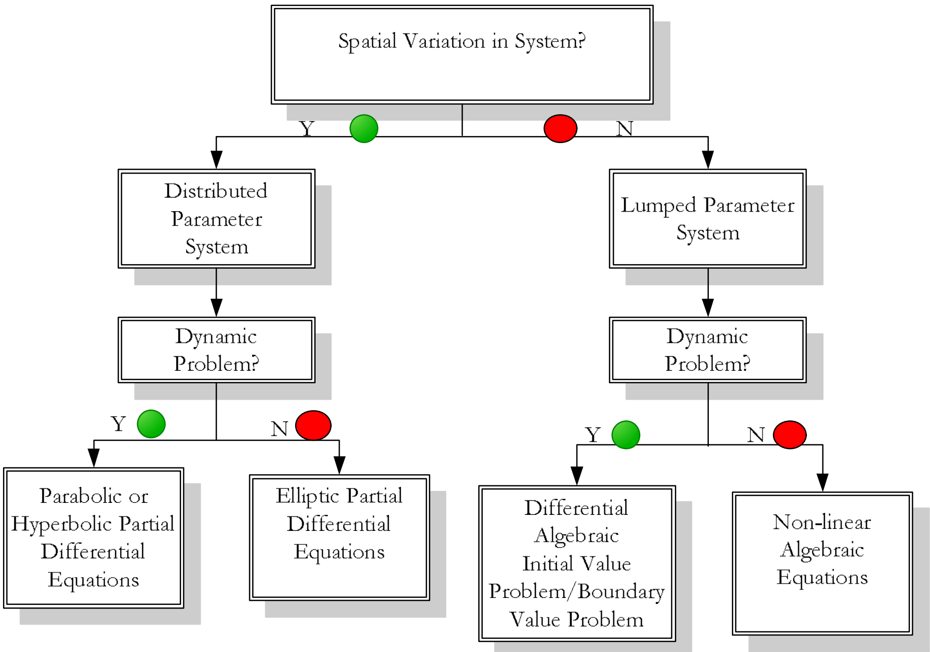


Figure 23: Classification of quantitative model characteristics. Adapted from Iyun(2005)

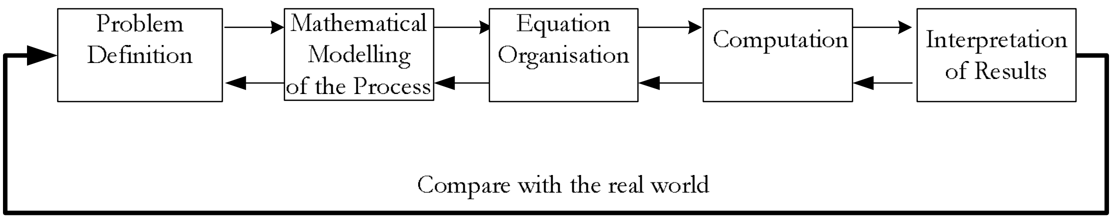


Figure 24: Quantitative process modelling strategy

Extended Kalman Filtering (EKF):

Kalman filter is a state estimation algorithm typically for linear dynamic system. The method has been modified in various forms under Extended Kalman Filtering (EKF) to suit various applications. For example, Villez *et al*., (2011) applied EKF in fault detection and identification in chemical process system to account for process non-linearities. The authors observed that EKF provides superior performance over traditional Kalman Filtering when dealing with non-linearity in state estimation of chemical processes. The algorithm employs local linearization approximation of the non-linear system under consideration. The states of the system are considered as a set of points that are deterministically selected from an approximate Gaussian distribution. Several authors including Farrell & Ioannou (2001), Julier, *et al*.,(2000) and Ricker & Lee (1995) have also commented on the adaptations to EKF to suit various application areas and to enhance EKF algorithm efficiency.

Particle Filtering

Kalman Filters and their extended versions assume that the posterior probability density at every time step is Gaussian and hence, parameterized by a mean and variance. For non-linear/non-Gaussian state estimation problems, particle filtering techniques have been successfully applied (Prakash, *et al*., 2011). Particle filters are sequential Monte Carlo methods based on point mass ("particle") representations of probability densities and require a proposal distribution. The choice of proposal distribution is the fundamental design issue in the application of particle filters (Arulampalam, *et al*., 2002; Prakash, *et al*., 2011)

Shortcoming of Model-Based Quantitative Approach:

The complexity of modern engineered systems, such as a large petrochemical refinery, with superficial human knowledge of the resulting complex systems hide the necessary details needed to build a detailed quantitative model and in situations where complex models result from complex systems, the required computational resources are non- trivial. The scope and depth of the purpose for a model will determine the ultimate complexity of the final mathematical description.

One way to get around the constraint of incomplete or uncertain knowledge of complex systems, without using purely qualitative methods, is the use of qualitative physics-based techniques (De Kleer and Brown, 1984; Kuipers, 1986; Sharma, 1997). The key advantage of qualitative physics-based models is that they enable conclusions about a process without exact expressions governing the process and precise numerical inputs. In some cases, partial conclusions can be reached from incomplete and uncertain knowledge of the system under consideration. Fault detection and diagnosis (FDD) based on quantitative models is unlikely to emerge as the method of choice in the near future because of the weaknesses pointed above, but simplified physical models will continue to make inroads into FDD applications (Katipamula and Brambley, 2005).

2.3.2 Qualitative Model-Based Approach

Representing a physical process qualitatively does not involve detailed mathematics but provides useful insight into the fundamental understanding of the process under consideration, such as a large chemical plant. Qualitative approaches to modelling, such those based on logic and qualitative states (De Kleer and Brown, 1984; Forbus, 1993; Kuipers, 1986) have the advantage of easy understanding, reduced modelling efforts and computational resources but at the expense of exact diagnostic resolution when compared with their quantitative counterparts. One approach to classification of qualitative model-based methods is depicted hierarchically in Figure 25.
Causal qualitative model and its connectivity matrix representation are the backbone of the thesis. These concepts are discussed in details in Section 2.3.3.

2.3.3 Qualitative Causal Models

This section discusses causal qualitative approach to process modelling in detail due to its relevance to the thesis. It describes graph theory, a fundamental of the directed graph and connectivity matrix methods used in the thesis. The chapter extends discussions on graphs to usability of directed graphs with respect to process schematics and in the storage and processing of graphs representations in computer programs. This section also explains connectivity matrix and XML because of their fundamental importance to the thesis. The section concludes with an illustrative example.

Cause-and-Effect Analysis

Studies in most physical systems involve the determination of cause-and-effect relationships among variables making up the system or events taking place within the system. For example, an increase in temperature (cause) of a confined gaseous system under an ideal gas law and constant volume leads to an increase in pressure (effect). The two variables are said to be correlated.

Detection of correlation by itself does not allow causality to be inferred. However, analysis of the time trends of process measurements has additional information because of the time dimension. Hence causal correlations can be detected such as by looking for cases when the correlation is maximized if one time trend is time-shifted relative to another.

It is important to derive appropriate methodology for deriving such relationships from available information sources such as process data history and process plant connectivity information as contained in the process P&IDs.

Cause-and-effect analysis is a systematic and well-documented diagrammatic technique designed to unearth the root-cause of problems and subsequent effects (MindGenius, 2008). It shows the effect of one process variable on the other. Effective cause-and-effect analysis can create extremely valuable benefits to process plant operator and control engineers in gaining insight into process analysis and diagnosis of root-causes of distributed faults in a process plant.

Qualitative model-based approach

- Causal models
  - Qualitative physics
    - -de Kleer and Brown (1984)
    - -Kuipers (1986)
    - -Iwasaki and Simon(1986)
    - -Shama(1997)
  - Digraphs and signed digraphs
    - -Iri et al.(1979)
    - -Umeda et al.(1980)
    - -Shiozaki et al.(1985)
    - -Kramer and Palowitch(1987)
    - -Oyeleye and Kramer(1988)
    - -Mohindra and Clark(1993)
    - -Vedam and Venkatasubramanian(1997)
    - -Zhang et al.(2011)
  - Fault trees
    - -Lapp and Powers(1977)
    - -Bossche(1991)
    - -Kavcic and Juricic(2001)
  - Multi Flow Modelling
    - -Lind(1999)
    - -Jorgensen(1993)
    - -Ohman(2000)
    - -Petersen(2000)
    - -Ouyang et al.(2005)
  - Flowsheet modelling
    - -Waschler et al.(2002)
    - -Maurya et al.(2004)
    - -d'Anterroches and Gani(2004)
    - -Ishii and Otto(2008)
    - -Walker(2009)
  - CAEX (XML-based and IEC/PAS 62424 compliant)
    - -Fedai and Drath(2005)
    - -Scherf(2006)
    - -Yim et al.(2006)
    - -Beez et al.(2008)
    - -Thambirajah et al.(2009)
  - Adjacency / Connectivity Matrix
    - -Yim et al.(2006)
    - -Thambirajah et al.(2009)
    - -Jang et al.(2009)
  - Intelligent P&ID (XML- based and ISO 15926 compliant)
    - -Alabi et al.(2011)
- Rule-based
  - Expert systems
  - First principle based
  - Limit and alarm

Figure 25: Qualitative model-based classification. The research remit is located at the lower levels of the hierarchy on the left hand branch

Causation is a relation between particular events: something happens and causes something else to happen. An event *A* can have more than one cause, none of which alone suffice to produce *A*. An event *A* can also be over determined: it can have more than one set of causes that suffices for *A* to occur. It is assumed (Spirtes, 2010) that causation is (usually)

(i) transitive
(ii) irreflective, and
(iii) antisymmetric.

This implies that:

- if *A* is a cause of *B* is a cause of *C*, then *A* is a cause of *C*
- event *A* can not cause itself, and
- if *A* is a cause of *B*, then *B* is not a cause of *A*

Causal modelling can be useful to the plant operator in the following ways:
Firstly, it provides a structured, systematic, methodical approach, ensuring that no important cause is overlooked. Secondly, it focuses on identifying contributing factors and causal effect to the problem, assuring that all causes are identified and the root-cause of the fault is identified so that appropriate action can be taken. Thirdly, it allows visual identification of possible causes, implying that once the plant operator sees the causes laid out, more thoughts are triggered and gaps are identified. Diagnostic problems infer system malfunctions from observables using abductive reasoning, which is a process of generating a plausible explanation for a given set of observations of facts.

2.3.4 Graphs

One way of representing physical systems qualitatively is the use of graphs. The use of directed graph (DG or "digraph") and signed directed graph (SDG), in particular, is a well researched and accepted approach of representing causality and finds application in fault detection, propagation and diagnosis(Gao, *et al*., 2010). For instance, DG can be used to generate connectivity matrix of various plant items, connections and direction within a process plant (Jiang, *et al*., 2009; Thambirajah, *et al*., 2009). Causal search techniques can be used to trace process malfunctions to their root source.

2.3.5 Fundamentals of Graph and Digraphs

This sub-section provides an introduction to basic graph theory in order to establish its relevance to the current work. The subject of graph theory has a very strong mathematical foundation which has been studied for and has been successfully applied in process engineering practice to estimate cause-and-effect relationships among process variables and relationships among plant items (Maurya, *et al*., 2003). Due to different terminologies and denotations used to describe graphs and graphs representations, the terminologies used in this thesis closely follow those used by Gross and Yellen (2004).

Basic Terminology of Graph Theory

This section presents relevant terminologies used in describing a graph and its properties such as those shown in Figure 26.

D1: A graph $G = (V, E)$ consists of an ordered pair (V, E)

- The elements of V are called the vertices
- The members of E called the edges, which are pairs of vertices (an ordered pair in a directed graph and unordered in an undirected graph).
- Each edge has a set of one or two vertices associated to it, called its endpoints. An edge is said to join its endpoints.

D2: If vertex v is an endpoint of edge e, then v is said to be incident on e, and e is incident on v

D3: A vertex u is adjacent to vertex v if they are joined by an edge.

D4: Two adjacent vertices may be called neighbours.

D5: Adjacent edges are two edges that have an endpoint in common.

D6: A proper edge is an edge that joins two distinct vertices.

D7: A multi-edge is a collection of two or more edges having identical endpoints.

D8: A simple adjacency between vertices occurs when there is exactly one edge between them.

D9: The edge-multiplicity between a pair of vertices u and v is the number of edges between them.

D10: A self-loop is an edge that joins a single endpoint to itself.

D10: A simple graph is a graph that has no self-loop or multi-edges.

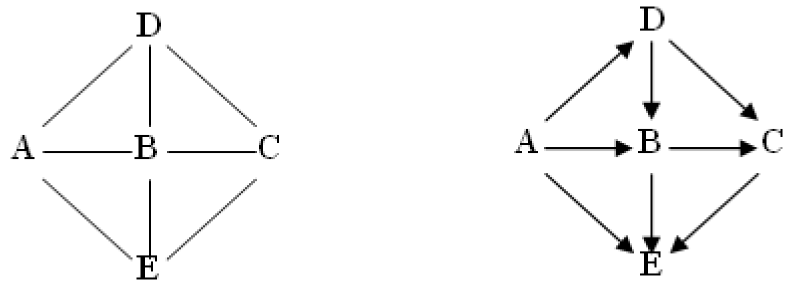
In signed directed graphs, the path between two vertices is assigned sign ‘‘+” if it represents positive influence (reinforcement) and sign “-” if it represents negative influence (suppression).

Illustration of Fundamentals of Graph Theory

A graph is an ordered pair <*V*, *E* >, where *V* is a set of vertices, and *E* is a set of edges. The members of *E* are pairs of vertices (an ordered pair in a directed graph and unordered in an undirected graph). For example, the edge *A* → *B* is represented by the ordered pair <*A*, *B* >.

In directed graphs, the ordering of the pair of vertices representing an edge in effect marks an arrowhead at one end of the edge. Formally, a graph is an ordered triple <*V*, *M*, *E* > where *V* is a non-empty set of vertices, *M* is a non-empty set of marks , and *E* is a set of sets of ordered pairs of the form *{*[*V1*, *M1*], [*V2*, *M2*]*}*, where *V1* and *V2* are in *V*, $V1 \neq V2$, and *M1* and *M2* are in *M.*

For illustrations, consider Figure 26, undirected (i) and directed (ii) graphs respectively,



(i) Undirected graph-contains only undirected edges.(e.g. *A* - *B*)

(ii) A directed graph, containing only directed edges. (e.g. *A* → *B*)

Figure 26: Schematic representation of (i) undirected and (ii) directed graphs

If *G* = <*V*, *M*, *E* >, *G* is said to be over *V*. From the right panel (digraph) in Figure 26, *G* can be represented as:

G = < {*A*, *B*, *C*, *D*, *E*}, {*EM*, >}, {{[*A*, *EM*], [*B*,>]}, {[*A*, *EM*], [*E*,>]}, {[*A*, *EM*], [*D* >]}, {[*D*, *EM*], [*B* >]}, {[*D*, *EM*], [*C*, >]}, {[*B, EM*], [*E*, >]}, {[*B*, *EM*], [*C*, >]}, {[*C*, *EM*], [*E*, >]}}>

Each member {[*V1*, *M1*], [*V2*, *M2*]} of *E* is an edge and represents the number of connections among the various nodes or vertices.

Each vertex *V1* of edge {[*V1*, *M1*], [*V2*, *M2*]} is called an end point of the edge. (E.g. *A* is an endpoint of {[*A*, *EM*], [*B*,>]}). *V1* and *V2* are adjacent in *G* if and only if there is an edge in *E* with endpoints *V1* and *V2*. E.g. *A* and *B* are adjacent, but *A* and *C* are not.

Undirected graph implies that set of marks *M* = {*EM*}, whereas, directed graph has *M* = {*EM*, >}, and for each edge in *E*, one edge-end has mark *"EM"* and the other edge-end has mark ">". For example, an edge {[*A*, *EM*], [*B*,>]} is a directed edge from source *A*, to sink *B* of the path. An edge {[*A*, *M1*], [*B*,>]}, is into *B* while and edge {[*A*, *EM*], [*B, M2*]} is out of *A*.

If there is a directed edge from *A* to *B*, then *A* is a parent of *B* and *B* is a child of *A*. *Indegree* of a vertex *V* is equal to the number of its parents and *outdegree* is equal to the number of its children and the degree is equal to the number of vertices adjacent to *V*. This is equal to the sum of *indegree* and *outdegree* in a directed graph. For instance, in the right hand panel of Figure

26, *B*'s parents are *A* and *D* and *B*'s children are *C* and *E*. Hence, *B* is of *indegree* 2, *outdegree* 2, and therefore, *degree* 4.

Sets of edges of path <*A*, *B*, *C*, *D* > will be {[*A*, *EM*], [*B*,>]}, {[*B*, *EM*], [*C*, >]}, {[*C*,>], [*D*, *EM*]}.
An edge {[*X*, *M1*], [*Y*, *M2*]} is in path *U* if and only if *X* and *Y* are adjacent to each other (in either order) in *U.*

Acyclic path contains no vertex more than once else it is cyclic. Intersection of two paths occurs if they have a vertex in common, called a point of intersection.

A subgraph of <*V*, *M*, *E* > is any graph <*V'*, *M'*, *E'* > such that *V'* is included in *V*, *M'* is included in *M* and *E'* is included in *E*. Figure 27 are subgraphs of Figure 26 (right hand panel)

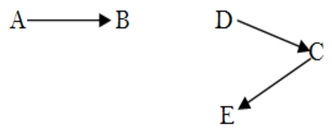


Figure 27: Decomposition of directed graph in Figure 26 into two sub graphs

In signed directed graphs, the path between two vertices is assigned sign "+" if it represents positive influence (reinforcement) and sign "-" if it represents negative influence (suppression).

<u>*Bond Graphs (BG)*</u>

Another type of graph, known as a *Bond Graph* (Beez, *et al*., 2008; Bouamama, *et al*., 2006; Bouamama, *et al*., 1997) provides a domain-independent topological representation that captures energy-based interactions among the different physical processes that make up the system. The vertices in the graph represent subsystems modelled as generic physical processes.

2.3.6 Application of Graph Theory in Process Analysis and Fault Diagnosis

Several researchers (Gao, *et al*., 2010; Kramer and Palowitch, 1987; Wakeman, *et al*., 1997) have used directed graphs for other diagnostic purposes. The goal of the diagnosis methods using signed directed graph (SDG) is to determine the fault set (Mohindra & Clark, 1992). It is imperative, however, that true process faults must be an element of the fault set provided the SDG model is complete. The fault sets will, however, probably contain extra elements since it provides all the possible explanations. Shiozaki *et al*., (1985) utilized a SDG in a fault diagnostic algorithm. However, the time complexity of the algorithm during implementation was very high. Kramer & Palowitch Jr. (1987) converted a SDG representing a physical process into a concise set of logical rules which provided a framework for addressing the issue of improved diagnostic resolution and reduced computational time. Processing time for a SDG comprising 99 nodes and 207 branches in Shiozaki *et al*., (1985) work took five minutes. However, with the technique of Kramer & Palowitch Jr. (1987), the processing time was reduced to a few seconds on a similar machine.

Directed graphs, such as the one shown in the left hand panel of Figure 28 can be utilized in the generation of the connectivity matrix, right hand panel of Figure 28 of a process plant.

2.3.7 Connectivity Matrix from Directed Graphs

The left hand panel in Figure 28 depicts a directed graph and its corresponding, on the right hand panel, connectivity matrix. The intersection of two nodes (row, column) denoted by “1” represents connection between the nodes while a “0” indicates no connection. One of the key objectives of the PhD work is to generate the connectivity matrix automatically from process schematic such as P&ID. The next subsection discusses connection matrices because of their importance to the thesis.

Connection Matrices

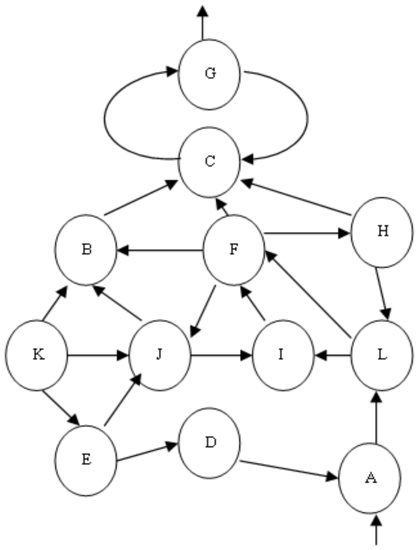
Most process structures can be reduced to interconnections of arcs and nodes as contained in P&IDs and illustrated mathematically by a matrix. The allocation of nodes and arcs to the rows and columns of the resulting matrix determines the type of matrix produced (Hartmann & Kaplick, 1990). The following types of connection matrix can be identified:

Connectivity matrix: the nodes are assigned to the rows and columns of the matrix.

Arc adjacency matrix: the arcs are assigned to the rows and columns.

Incidence matrix: the arcs are assigned to the rows and nodes to the column.

In the variants of connection matrices above, the following premises hold:



Connectivity matrix, A =

	A	B	C	D	E	F	G	H	I	J	K	L
A	0	0	0	0	0	0	0	0	0	0	0	1
B	0	0	1	0	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	1	0	0	0	0	0
D	1	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	1	0	0	0	0	0	1	0	0
F	0	1	1	0	0	0	0	1	0	1	0	0
G	0	0	1	0	0	0	0	0	0	0	0	0
H	0	0	1	0	0	0	0	0	0	0	0	1
I	0	0	0	0	0	1	0	0	0	0	0	0
J	0	1	0	0	0	0	0	0	1	0	0	0
K	0	1	0	0	1	0	0	0	0	1	0	0
L	0	0	0	0	0	1	0	0	1	0	0	0

Figure 28: Directed graph and corresponding connectivity matrix

$$\text{Connectivity matrix, } C_{ij} = \begin{cases} 1 \text{ if there is an arc from node } x_i \text{ to node } x_j \\ 0 \text{ if there is no arc from node } x_i \text{ to node } x_j \end{cases}$$

$$\text{Arc adjacency matrix, } A_{ij} = \begin{cases} 1 \text{ if there is a node connecting arc } u_i \text{ to arc } u_j \\ 0 \text{ if there is no node connecting arc } u_i \text{ to arc } u_j \end{cases}$$

$$\text{Incidence matrix, } I_{ij} = \begin{cases} -1 \text{ if arc } u_j \text{ springs from node } x_i \\ 1 \text{ if arc } u_j \text{ ends in node } x_i \\ 0 \text{ if arc } u_j \text{ is not connected to node } x_i \end{cases}$$

Where,

node $X = (x_1 \ldots x_n)$ and arc $U = (u_1 \ldots u_m)$

2.3.8 Reachability Matrix

If A represents the connectivity matrix containing N nodes and K, the sum of the successive powers of A up to the power of N, the *reachability matrix, R,* due to Mah(1983) and demonstrated with adjacency matrix by Jiang, *et al*., (2009) is defined as follows:

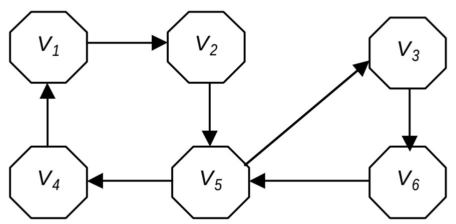
$$K = A + A^2 + A^3 + \ldots + A^N$$

The power to which A^p is raised indicates the number of p-step edge sequence traversed in moving from element i to j in matrix A^p.

$$R = (K)^{\#} = \begin{cases} \text{if } K(_{i,j}) = 0;\ R(_{i,j}) = 0 \\ \text{if } K(_{i,j}) \neq 0;\ R(_{i,j}) = 1 \end{cases}$$

As an illustrative example, consider the simple directed graph G shown and the corresponding connectivity matrix M depicted in Figure 29

$G =$



$M =$

	V_1	V_2	V_3	V_4	V_5	V_6
V_1	0	1	0	0	0	0
V_2	0	0	0	0	1	0
V_3	0	0	0	0	0	1
V_4	1	0	0	0	0	0
V_5	0	0	1	1	0	0
V_6	0	0	0	0	1	0

Figure 29: Simple directed graph with corresponding connectivity matrix

Matrices representing successive powers of M are shown in Figure 30-M^2, M^3, M^4, M^5, M^6. K is the sum of M through to M^6 and R is the reachability matrix of M.

	V_1	V_2	V_3	V_4	V_5	V_6
V_1	0	0	0	0	1	0
V_2	0	0	1	1	0	0
V_3	0	0	0	0	1	0
V_4	0	1	0	0	0	0
V_5	1	0	0	0	0	1
V_6	0	0	1	1	0	0

M^2

	V_1	V_2	V_3	V_4	V_5	V_6
V_1	0	0	1	1	0	0
V_2	1	0	0	0	0	1
V_3	0	0	1	1	0	0
V_4	0	0	0	0	1	0
V_5	0	1	0	0	1	0
V_6	1	0	0	0	0	1

M^3

	V_1	V_2	V_3	V_4	V_5	V_6
V_1	1	0	0	0	0	1
V_2	0	1	0	0	1	0
V_3	1	0	0	0	0	1
V_4	0	0	1	1	0	0
V_5	0	0	1	1	1	0
V_6	0	1	0	0	1	0

M^4

	V_1	V_2	V_3	V_4	V_5	V_6
V_1	0	1	0	0	1	0
V_2	0	0	1	1	1	0
V_3	0	1	0	0	1	0
V_4	1	0	0	0	0	1
V_5	1	0	1	1	0	1
V_6	0	0	1	1	1	0

M^5

	V_1	V_2	V_3	V_4	V_5	V_6
V_1	0	0	1	1	1	0
V_2	1	0	1	1	0	1
V_3	0	0	1	1	1	0
V_4	0	1	0	0	1	0
V_5	1	1	0	0	1	1
V_6	1	0	1	1	0	1

M^6

$K = M+M^2+M^3+M^4+M^5+M^6 \quad =$

	V_1	V_2	V_3	V_4	V_5	V_6
V_1	1	2	2	2	3	1
V_2	2	1	3	3	3	2
V_3	1	1	2	2	3	2
V_4	2	2	1	1	2	1
V_5	3	2	3	3	3	3
V_6	2	1	3	3	3	2

$$R = (K)^{\#} = \begin{cases} \text{if } K(_{i,j}) = 0;\ R(_{i,j}) = 0 \\ \text{if } K(_{i,j}) \neq 0;\ R(_{i,j}) = 1 \end{cases}$$

	V_1	V_2	V_3	V_4	V_5	V_6
V_1	1	1	1	1	1	1
V_2	1	1	1	1	1	1
V_3	1	1	1	1	1	1
V_4	1	1	1	1	1	1
V_5	1	1	1	1	1	1
V_6	1	1	1	1	1	1

Figure 30: Iterative steps for transforming connectivity matrix into reachability matrix

The *reachability matrix* provides a method of determining which elements can be reached from a particular starting point by examining the non-zero row entries. For the illustrative example in Figure 30, the recheability matrix *R* has non-zero entry indicating that every node can be reached from any starting node. Using the *reachability matrix*, it is possible to trace which elements are connected and how many steps or elements lay along the path. For example, M^3 in Figure 30 suggests that there are three-step edge sequences from:

V_1 to V_3 ($V_1 \rightarrow V_2 \rightarrow V_5 \rightarrow V_3$)
V_1 to V_4 ($V_1 \rightarrow V_2 \rightarrow V_5 \rightarrow V_4$)
V_2 to V_6 ($V_2 \rightarrow V_5 \rightarrow V_3 \rightarrow V_6$)
V_5 to V_5 ($V_5 \rightarrow V_3 \rightarrow V_6 \rightarrow V_5$) - a loop

The practical application is that it is possible to do a quick check on items of equipment along a path by specifying the starting element and thus give an insight into possible propagation of disturbance along the direction of process fluid and signal flow.

The concept of reachability matrix will be applied to the Tennessee Eastman case study in Section 5.7 to give an insight into the process.

2.3.9 Comparison of Full Process Model with Connectivity Model for Fault Diagnosis

Table 2 compares the effectiveness of full process models, such as physical and empirical models with connectivity model generated from connectivity and directionality information such as contained in process P&ID. A "√" indicates strength while "X" denotes shortcoming.

It can be concluded from the comparisons in Table 2 that there are merits and disadvantages inherent in either of the approaches.

Table 2: Comparison of relative strengths and weaknesses of full process model against connectivity model for fault diagnosis. √ indicates strength and X indicates weakness

Requirement	Full process model	Connectivity model
Suitability	√	X
Robustness	X	√
Human interpretation	X	√
Quick detection and diagnosis	X	√
Isolability	X	√
Novelty identifiability	√	X
Error classification	√	X
Adaptability	X	√
Explanation facility	√	√
Modelling requirement	X	√
Storage and computational requirements	X	√
Multiple fault identification	√	X

The graph- based methods discussed above seem very powerful as shown by Jiang, *et al*., (2009) for example. However, their application is limited at present by the bottleneck of creating the graph or connectivity matrix from the drawing. Generation of connectivity matrix in an automated fashion and on a very large scale from readily available electronic process schematics has been accomplished in this thesis.

2.3.10 Manipulation of Connectivity Matrix

The next task after extraction and storage of items of interest and their connectivity information from process schematic is to manipulate the connectivity matrix created. This section describes established algorithms for searching graphs and graph related objects. A modified version of traditional graph traversal technique used in this thesis is presented.

Depth First Search: The depth-first search approach utilizes last in first out (LIFO) data structure. The starting element is put on a stack while an adjacent element not yet visited is explored continuously until the last item has been visited. The search then backtracks to the previous element visited to explore any unvisited element. An implementation of depth first search algorithm due to Thambirajah *et al*., (2009) used in this thesis is depicted in Figure 31.The figure shows the search from *Element-006* to *Element-001*..

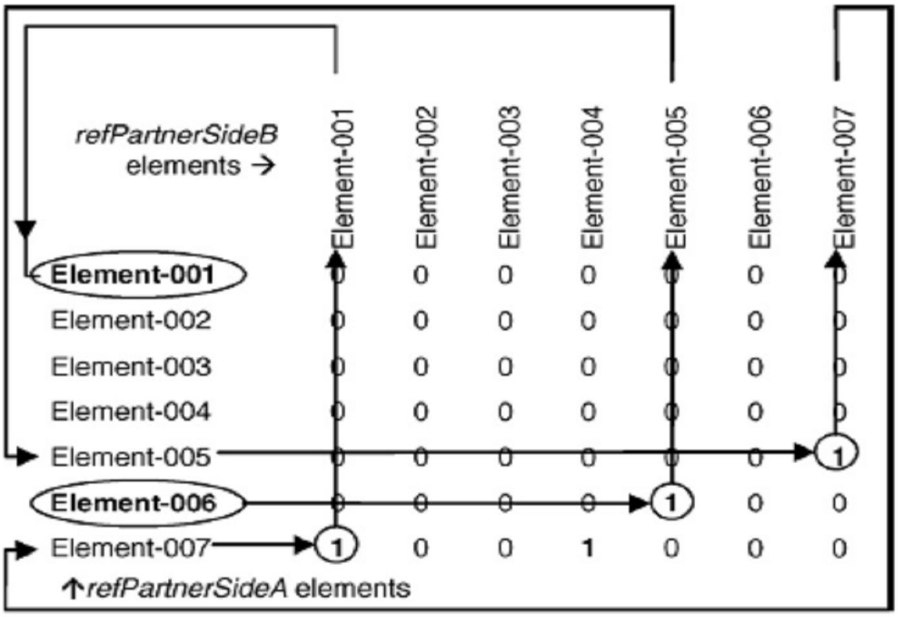


Figure 31: Depth-first search used to compile a list of forward path elements (From Thambirajah *et al*., 2009)

All the elements connected to a starting element on the first column is depicted by a "1" entry in the connectivity matrix.

Breadth First Search: This is an alternative search strategy to the depth first approach. Here, the graph traversal makes use of queue data structure – first in first out (FIFO). The starting element in the array of items to be searched is put on the queue while all elements connected or adjacent to the first element is visited. Each element visited is added to the queue as it is visited until all elements have been exhausted. Each element on the queue is visited to explore its adjacent elements until all items on the queue have been visited

2.3.11 eXtensible Mark-up Language (XML)

One of the aims of the thesis is to manipulate connectivity matrices for the purposes of tracing the root-cause of a disturbance along the direction of process fluid and information flow with a view to confirming the root source of the disturbance and eliminating spurious results from data-driven analysis. Consequently, it is important to pay attention to the automated generation of such matrices from available information sources such as a PFD or P&ID.

This section describes XML, a text-based platform and vendor independent data description for storage, transmission and integration of data from a variety of sources. Plant description using the XML provides greater flexibility in computer manipulation of process schematics when compared with the graphic format.

Hence, for any engineering drawing such as a PFD or P&ID to be amenable to algorithmic manipulation in a computer program, such drawings must be converted and represented electronically in a format that a computer program can read and operate on. One of such useful computer representation of engineering drawings is the XML.

The following subsections explain the structure of an XML file because XML is used extensively and feature prominently in the thesis.

XML is a system and hardware independent language for expressing data and its structures within an XML document and conforms to ISO/IEC standardization (w3.org, 2011). An XML document is a text file that contains the data together with mark-up that defines the structure of the data for easy data communication from one computer to another.

An XML document basically consists of two parts, a *prolog* and a *document body.* The prolog provides information necessary for the interpretation of the contents of the document body. It contains two optional components, and since both components can be omitted, the prolog itself is optional. The two components in the sequence that must appear are: *XML declaration*, which defines the version of the XML and may specify the particular *unicode* character encoding used in the document; *document type declaration* (DTD), which specifies an external DTD that identifies mark-up declarations for the elements used in the body of the document, or explicit mark-up declaration, or both. The *document body* contains the data. It comprises one or more elements where each element is defined by a begin tag and end tag. The elements in the document body define the structure of the data. There is always a single root element that contains all other elements.

An XML document must be *well-formed* and *valid* before being processed by an *XML processor,* which may be validating or non-validating. A *well-formed* XML document implies that the document conforms to the rules for writing XML, as defined by the XML specification. A valid document is a well-formed document that has an associated DTD. A DTD essentially defines a mark-up language for a given type of document and is identified in the *DOCTYPE* declaration in the document prolog. Below is an example of a well-formed and valid XML document for describing a process plant.

```
<?xml version="1.0" encoding="UTF-8" standalone=”yes”?>

<!DOCTYPE Plant  SYSTEM “ProcessPlant.dtd ”>

<Plant PlantID= “Reformer101” >

</Plant>
```

The document consists of a root element that defines the plant with a prolog above the root element. The value of “yes” assigned to *standalone* implies that the document is independent on any external definition of mark-up, which means the document is self-contained. The default value of *standalone* is “no”, so this can always be left out when a document is not standalone. The name that appears in the *DOCTYPE* declaration, in this case *Plant* must always match that of the root element for the document. XML mark-up divides the contents of a document up into *elements,* Plant in this simple example, by enclosing segments of the data between tags, a start tag and an end tag. Document comments can go anywhere in the prolog or document body, but not inside a start tag or an end tag or within an empty tag, for instance, a comment might go like this:

```
<variable temperature= “120” pressure= “48” flowrate= “160” ></variable>

<!--This would normally be written in the shorthand form, like this:  -->

< variable temperature= “120” pressure= “48” flowrate= “160”  />
```

Additional information can be put within an element in the form of one or more attributes. An attribute is defined by an attribute name, and the value is specified as a string between the quotes as:

```
<elementname attribute= “Attribute value”>

</elementname>
```

2.4 Process History-Based Methods

Advancements in computer data processing capabilities coupled with ever decreasing costs of hardware have taken computer data processing capabilities and extraction of useful information from measured process data to a higher level.

In contrast to model-based approaches where *a priori* knowledge (known or assumed ahead of time) about the model (quantitative, qualitative or both) is assumed; in process history-based methods, only the availability of a large amount of historical process data is assumed (Venkatasubramanian, *et al*., 2003b). When large quantities of data generated from process measurements are analyzed using algorithms such as the PCA technique to reduce the dimensionality of the data, it produces valuable information and insights regarding the state of the process under consideration. It can also be used to extract certain features about the process which can be fed as *a priori* knowledge to a diagnostic tool. A hierarchical classification of some methods for data-driven analysis based on (Thornhill and Horch, 2007) is presented in Figure 32. The classification considers process history based analysis under two broad headings-qualitative and quantitative methods.

In process history-based or data-driven methods, both process inputs and outputs are measured at regular intervals called sampling time. The sampled data are either used instantly in real-time operations or archived for process monitoring and auditing. Several researchers have exploited process data history to perform FDD and to find direction of fault propagation (Bauer and Thornhill, 2008; Thornhill and Horch, 2007).

Process History-based Fault Detection

The use of process measurements to detect process anomalies is an active research area and the various techniques developed have been successfully deployed in the industry. For example, AlGhazzawi and Lennox (2008) described a real-time, recursive multivariate statistical analysis based on PCA to detect sub-optimal process performance. The technique was able to capture dynamic characteristics of the process, a weakness in static PCA models.

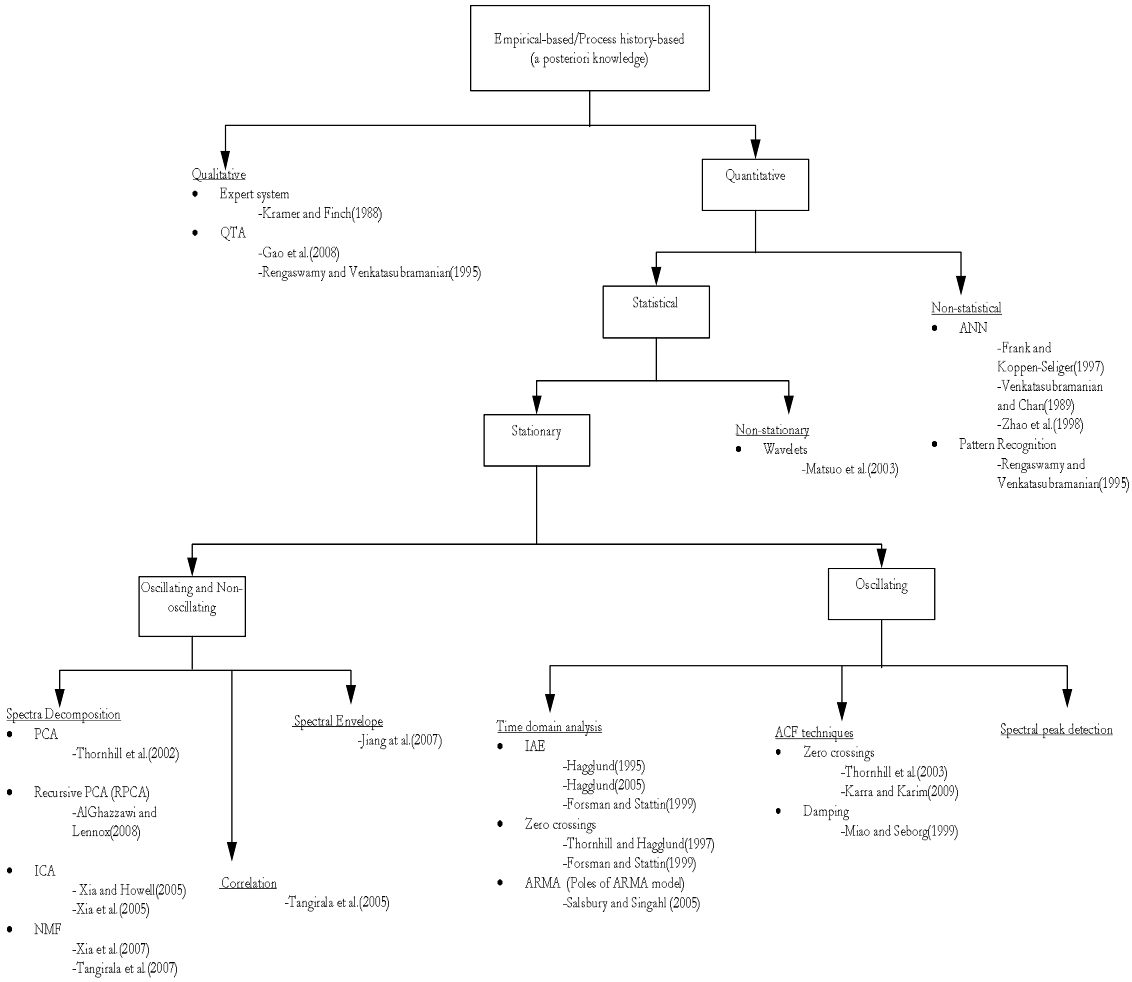


Figure 32: Classification of methods for data-driven plant-wide disturbance based on Thornhill & Horch (2007)

The classification in the lower layer of Figure 32 groups plant-wide disturbances into those oscillating, non-oscillating and non-stationary with the methods for detecting each classification of disturbance. However, experience has shown that the most commonly encountered form of disturbance in process industry is that of oscillation.

In an attempt to deal with the problems of measurement noise, missing values, outliers and time delays in measurements used for signal analysis, Thornhill, *et al.,* (2002) used spectral analysis of process signals in their study. The signal decomposition analysis was carried out by subjecting the signals in the time domain to Fourier analysis resulting in frequency domain analysis. The resulting multivariate analysis was carried out in the frequency domain. The steps involved in the procedure are summarized and depicted in Figure 33.

The reason for explaining the algorithm in detail is that the methods have been implemented in the signal processing tool, plant disturbance analysis (PDA®) used in process measurements analysis in this thesis, hence fundamental understanding of the algorithm is essential.

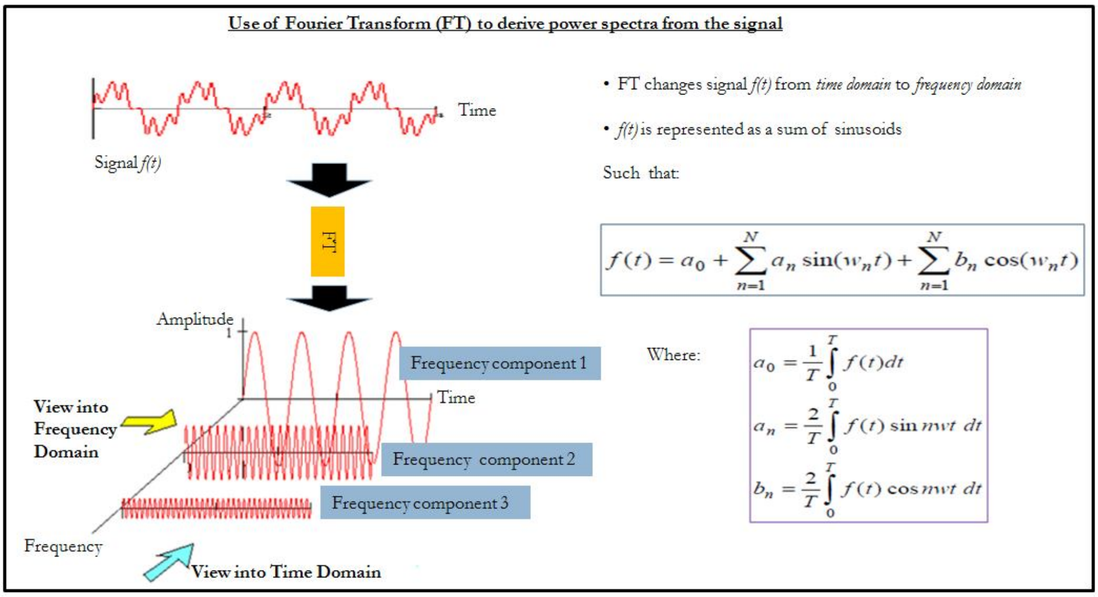


Figure 33: Signal decomposition to constituent sinusoids using a Fourier Transform

Fast Fourier algorithm (FFT) is used to carry out efficient signal transforms from time domain to frequency domain. The result of FFT application is summarized in Figure 34 which shows the frequency components of the original signal.

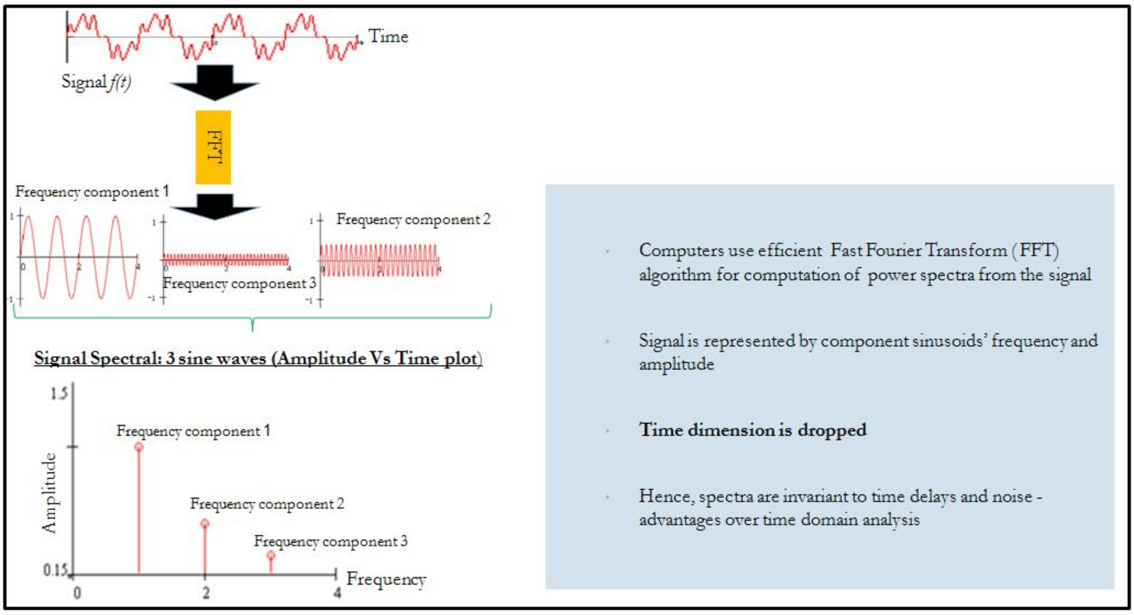


Figure 34: Output from Fast Fourier Transform of process signal

The constituent sinusoids resulting from signal decomposition shows the amplitude on the vertical axis and frequency on the horizontal axis. The time component of the original signal has been eliminated, hence, the analysis is unaffected by time delays and noise which are advantageous over time domain analysis.

Each process measurement was decomposed into its respective frequency spectrum up to the Nyquist frequency (one-half of the sampling frequency) which was used to create a matrix, *X* whose rows are the single-sided power spectra of the signals.

PCA analysis decomposes *X* as a sum of orthogonal basis functions, *W*, called the loadings, ranked according to weightings, *T*, called the scores. X can be rewritten as $X = TW$ as illustrated in Figure 35.

In the analysis carried out by Thornhill, *et al*., (2002), a three principal components model (three basis functions, *w*) captured 90% of the variance and the plot in 3-D plot and showed clusters with each cluster containing possible root-cause(s) of oscillation within the process.

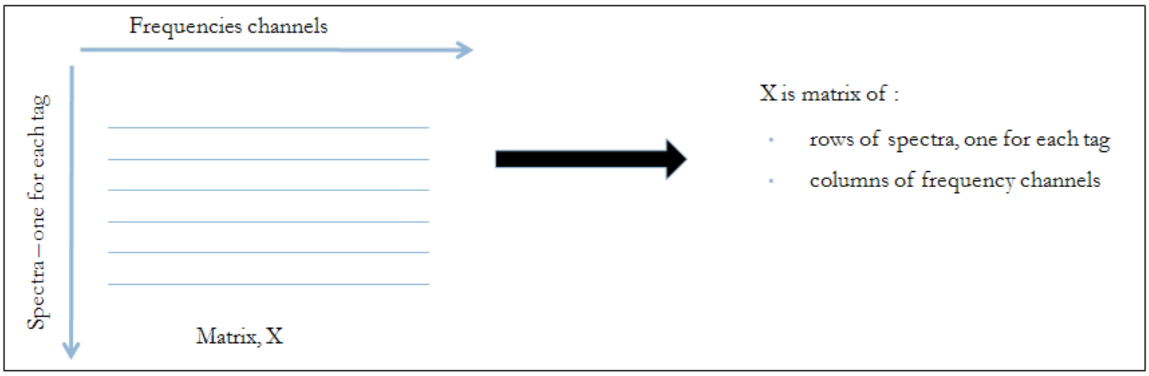


Figure 35: Signal decomposition and transformation into spectra for PCA analysis

<u>Process History-Based Fault Diagnosis</u>

Thornhill *et al*., (2003) used measured process data to derive a numerical non-linearity index which was subsequently used to detect a disturbance that propagated plant-wide and to identify the root-cause of a process disturbance. Non linearity is strongest at the root source of the disturbance under consideration. The underlying concept in reaching a conclusion about root-cause in the work is based on the observation that non-linearity reduces due to the filtering nature of the process as the disturbance propagates away from the root-cause. Bauer

& Thornhill (2008) applied cross correlation function to estimate time delay between process measurements which was used to derive the propagation path in the form of a causal map. These works and others show the potential of applying process history data in FDD.

The relevance of the discussions above stemmed from the fact that the PDA tool used for finding data-driven root-cause hypothesis that are subsequently tested against the process topology's connectivity and directionality information captured in this thesis used spectral PCA algorithm calculations for data analysis.

2.5 Approaches to Fault Detection and Diagnosis (FDD)

Various paradigms have been used by several authors to implement FDD tools, such as those that combine knowledge about the process under consideration (model-based) with process measurements (model-free) from sensors and instruments. Irrespective of the technique adopted, it is important to be able to build robust tools capable of effectively diagnosing process upsets under various conditions coupled with reusable components that can be integrated with existing or new tool to enhance rapid and cost effective industrial FDD tool development (Struss, *et al*., 2010). Table 3 summarises some of the monitoring tools developed with a combination of the various methods.

Table 3: Some FDD methods and tools

FDD Method	FDD Tool	Faults Considered	Intended End User
FDD based on performance indices and expert rules	Performance validation tool	Poor controller tuning, faulty flow and temperature sensor, faulty actuators	Plant operators / maintenance personnel
Statistical analysis and minimization of mass and energy balance residual	Offline sensor validation	Bias and drift in temperature and flow rate sensors	Maintenance engineer, plant operators and commissioning engineers
FDD based on stochastic qualitative reasoning	Performance monitoring support system	Actuator, sensor and controller failures	Plant operators /maintenance personnel
FDD based on qualitative causal reasoning and sign directed graphs	Performance monitoring support system	Actuator, sensor and controller failures	Plant operator/maintenance personnel

FDD based on statistical analysis of residuals	Embedded performance monitoring system	Stuck damper	Plant operator
FD based on fault direction space method. FD using physical models and analysis of filtered residuals	Performance monitoring tool	Stuck or leaking heating and cooling coil valves, low heating steam supply temperature, reduced (increased) cooling water flow, incorrect flow rate	Plant operator
Qualitative model-based fault detection	Performance monitoring tool	Valve stuck, or with restricted range, sensor offset, excessive control signal	Plant operator
FDD using an expert system	Performance monitoring (audit) tool	Wrong pressure, simultaneous heating or cooling, defective sensor	Plant operator
Detection based on fuzzy expert rules and generic fuzzy models	Performance monitoring and automated commissioning tool	Leaky valve, fouled coil, valve stuck, open, midway or closed	Commissioning engineer, plant operator
FD based on statistical analysis of residuals	Performance validation tool	Blocked valve, stuck valve, partially open valve, faulty sensor	Maintenance personnel, process plant operator
FD based on fault-symptom tree expert rules	Performance monitoring system	High energy consumption(energy efficiency), poor control performance	Plant operator

One conclusion to be drawn from the list in Table 3 is that a combination of techniques is adopted for most FDD tools. The approach of combining techniques compensates for deficiencies in a single technique and thus produces superior results as demonstrated by (Chiang and Braatz, 2003; Iyun, 2005; Lee and Yoon, 2003).

Yim *et al*., (2006) described a tool developed by combining electronic description of the plant structure with results from signal-based analysis to isolate root-cause of plant-wide disturbance. In the work reported by Norvilas *et al*., (2000), multivariable statistical analysis was combined with a knowledge-based system for monitoring chemical process operation. Kosebalaban & Cinar (2001) employed MSPM, contribution plots, and parity space fault diagnosis techniques for detecting abnormal operation of dynamic processes and diagnosis of sensor and actuator faults. Thornhill *et al*., (Thornhill, 2005; 2003; Thornhill, *et al*., 2003) derived and used non-linearity index, zero crossings of auto covariance function (ACF) and spectral principal component analysis (PCA) from routine process measurements to detect and diagnose plant-wide oscillations. From the reviews, it is evident that a combination of techniques (hybrid) is widely employed and recommended in developing effective FDD tools.

Christofides *et al*., (2007) posited that a unified and effective approach to fault detection is elusive in practice because of changing conditions of data and plant operation. However, the various approaches to FDD in its most abstract form with some modifications in their implementation can be defined as a two-step task. The two steps can be considered as fault detection and identification or isolation.

- *Step 1.* Compares the actual behaviour of a process, as manifested by the values of the operating variables, against the behaviour predicted by a model or measurements, and generate the residuals which reflect the impact of faults. This is the fault detection stage.

- *Step 2.* Evaluates the residuals and through an inversion process, identify the inputs (i.e. faults) that caused the observed behaviour. The inversion process could be analytic or take on various forms of a decision process, such as hypothesis testing, logical testing against thresholds, pattern recognition (syntactic, or quantitative) and so on. This is the fault isolation step.

Methods for isolating faults include model-based and data-based methods. Model-based approaches use mathematical or logical representation of the process to design dynamic filters and compute residuals that cause specific faults. Data based methods on the other hand utilise measured data from the process variables to find and compare the location and direction of the system in the state-space with past behaviours.

Factors that are typically considered when applying the above generic approach to FDD include: sources of faults under consideration; failure modes to include for each source; types of models used to describe process behaviour; representation of process signals and computation of residuals which is usually defined by the type of process model used.

2.6 Hybrid (Model-Mixing) Approach

It can be emphasised that data, models, knowledge, and experience all have critical roles to play in effective fault isolation and management. This implies that any means of combining all these elements for better and improved diagnostic should be sought. This gives rise to the idea of model mixing or hybrid systems. Development of improved modelling paradigms that exploit process data and various forms of prior knowledge tends to produce a better diagnostic result. Iyun, (2005); Chiang & Braatz, (2003) and Lee, *et al*., (2003) among several other authors have concluded that the use of qualitative and quantitative techniques offers superior diagnostic performance.

2.6.1 Justification for Model-Mixing

Operations in process industries are typically non-linear. Interruptions and disturbances occur randomly and may move the process to a new operating point where the nonlinearity becomes apparent. Identifying and isolating faults plant-wide in the face of such nonlinearities, externally imposed variability and increasingly complex process plants becomes a challenge for process control engineers and plant operators using conventional FDD approaches(Chiang and Braatz, 2003). Novel, intelligent and robust techniques capable of coping with such plant complexities and non-linearity are needed.

Table 4 (Venkatasubramanian, *et al*., 2003c) summarises the effectiveness of various diagnostic methods in meeting expected goals. It can be inferred from the comparison table that no single technique meets all the desired criteria for fault diagnosis because no single method has all the desired properties. The conclusion to be drawn from the analysis is that a combination of methods stands a better chance of dealing more effectively with fault detection and isolation. Consequently, a hybrid approach or combination of techniques, therefore, looks very promising in overcoming inherent drawbacks in single FDD technique. This serves as a motivation for adopting a hybrid approach as used in the thesis to detect the root-cause of a plant-wide disturbance.

Table 4: Comparison of relative strengths and weaknesses of various diagnostic methods. The table indicates that no single approach satisfies all the diagnostic requirements (based on Venkatasubramanian, *et al*., 2003c)

Requirement	**Digraph**	**Expert system**	**PCA**	**NN**	**QTA**	**Observer**	**Abstraction hierarchy**
Quick detection and diagnosis	?	✓	✓	✓	✓	✓	?
Isolability	X	✓	✓	✓	✓	✓	X
Robustness	✓	✓	✓	✓	✓	✓	✓
Novelty identifiability	✓	X	✓	✓	?	?	✓
Error classification	X	X	X	X	X	X	X
Adaptability	✓	X	X	X	?	X	✓
Explanation facility	✓	✓	X	X	✓	X	✓
Modelling requirement	✓	✓	✓	✓	✓	?	✓
Storage and computation	?	✓	✓	✓	✓	✓	?
Multiple fault identification	✓	X	X	X	X	✓	✓

✓ = strength X = weakness ? = unknown

2.7 Fault Diagnosis in Process Industry

This section puts fault diagnosis into the context of process industry. Process plants used in the conversion of raw materials to finished products, such as a refinery, are typically non-linear and complex in nature, both in construction and operation. The plants generally make use of recycle streams and achieve energy efficiency through heat integration for maximum performance.

Process plants evolve over time and throughputs are dictated by prevailing economic and technological conditions. For example, a sudden demand on a particular product stream such as gasoline might require process and instrumentation modification to cope with the change. Similarly government regulation on effluent composition might change prior to discharge to the environment. The new requirements from existing process plant imply that modifications to the plant such as addition of instrumentation or modifications to operating mode become inevitable. The various changes can complicate root-causes of disturbances that have propagated plant-wide due to the complex nature of the resulting plant.

As shown in Table 5, possible sources of process plant upset can be broadly classified under four factors:

1. **Sources due to process operation:** this includes use of recycle streams, heat integration, fluctuations in raw material composition, start up and shut down operations.

2. **Faults emanating from process equipment:** process equipment has design capacity beyond which further output cannot be obtained from the asset. In some severe cases, attempts to operate beyond design capacity could lead to equipment breakdown and possible material release. An example is the maximum operating pressure of a reactor vessel. Other examples include valve saturation, equipment degradation due to wear and tear and equipment in poor state of repair or maintenance.

3. **Process upsets from control systems:** controllers rely on measurements taken from key process variables to take action. When such measurement data from sensors and other instruments are noisy, inaccurate or corrupt, controller performance is adversely

affected. Other factors that could cause a control system to exhibit and propagate process upset include poor controller parameter(s) (proportional integral derivative, PID) settings (tuning), controllers interacting with one another where output from one controller disturbs another controller, and valve fully opened or closed (saturation) and advanced control design implementation issues such as cascade control.

4. **Unknown or unpredictable sources:** these are the most difficult to deal with since neither the source nor the dynamic is known. It is usually difficult to model or predict such sources of upsets in real processes. There are numerous sources unknown process disturbances and few the sources include random changes in ambient operating conditions such as temperature and humidity, rain showers on process plant located in the open and thunderstorms.

Table 5: Causal factors in process plants. Sources of faults are broadly classified under four major headings: operation, equipment, control systems and unknown sources

Sources of faults in process plants			
Operation	**Equipment**	**Control Systems**	**Unknown**
• Use of recycle stream • Heat integration • Start up* • Shut down* • Change in raw material * special process operation	• Equipment operating at design limits such as valve saturation • Equipment degradation over time due to wear and tear • Equipment in poor state of repair	• Poor control structure design • Bad controller design • Noisy measurements • Poor controller tuning • Interactions among controllers • Advanced control strategy • Controller output at limits such as valve saturation	• Fluctuations in ambient temperature and humidity • Rain showers • Thunderstorm • Non-linearity • Fluctuations in power supply • Utility stability

2.8 Human Approach to Reasoning about Fault Diagnosis

Humans are capable of reasoning about their environment and subsequently draw conclusions in order to take an action without complete or accurate information about the environment. This is essentially similar to the process by which a computer system used in process automation takes control action viz: sensing, comparison with a reference and action.

However, humans unlike computer system do not require exact quantitative mathematical relationship about the world in order to reason and act accordingly. Humans are also capable of utilizing knowledge and information from a variety of sources to an overall image of the physical system under consideration. With time, humans are also able to learn from past experience and apply such knowledge in dealing with novel problems. Humans are highly flexible in their problem solving approach, are able to collaborate and communicate their findings to colleagues.

Despite the above listed qualities in humans' ability to reason about their environment in the face of insufficient data and uncertainty, an automated system is still preferable because humans suffer from a number of limitations. Humans' cognitive capability varies widely due to their levels of training and experience. When subjected to stressful conditions, repetitive tasks, large amount of data and information, humans' capability and capacity to reason degrades drastically. Similarly when experienced humans, with some fundamental knowledge of the process system leave a job or retire, all the experience and knowledge also goes with them. An automated diagnostic tool on the other hand can be used to train new operators and can cope with human limitations above-such as performing a boring or repetitive task.

2.8.1 Application of Artificial Intelligence (AI) Techniques and Expert System

This subsection introduces the application of AI in the research work, especially the knowledge-based system. AI techniques that have been successfully used and those with potential for use in FDD are also discussed. One advantage of AI techniques, such as artificial neural networks (ANN), over traditional approaches is the ability to offer novel solutions to problems devoid of a conventional solution.

The complex, non-linear and dynamic nature of engineered systems with incomplete knowledge about the resulting complex systems requires non-conventional approaches to control and monitoring. Output from researches in the field of AI, such as expert systems, neural networks and fuzzy logic have been able to help to some extent in coping with uncertainties and superficial knowledge about complex systems required for modern demanding and sophisticated systems. Expert systems, fuzzy logic and ANNs are some of the areas in the field of AI with increasing applications in complex systems whose problems are ill-defined or not thoroughly understood. These techniques can be used in diagnosing such complex systems as a chemical plant.

In process industries, the use of AI tools such as knowledge-based systems have continued to yield favourable results, especially in the area of control and optimization. Choosing the right AI tool for application in process industry is critical to the success of such AI approach and this approach in turn depends on the specific application. The right search algorithm should also be chosen. Forward and backward chaining search shown in Figure 36 are typically employed. For diagnostics tasks, such as utilized in expert systems, backward chaining searches are usually employed. This involves searching and reasoning from the conclusion (effect) backward, using sub-goals until the cause is discovered.

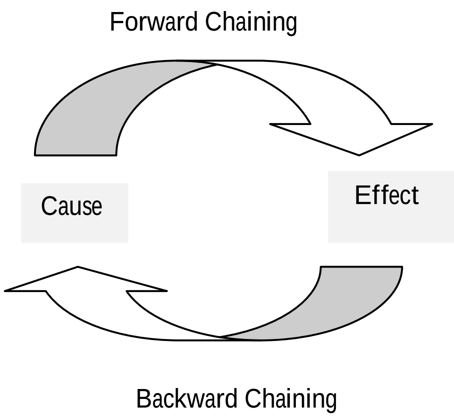


Figure 36: Search strategy for locating cause-and-effect in diagnostic tasks

Kokawa *et al*., (1983) used backward chaining technique to propose possible causes by working backward from all abnormal measurements to find common explanations. Prediction, on the other hand, uses forward chaining, reasoning from the known (cause) toward a solution. For example, Kramer & Palowitch (1987) propose possible root-causes for observed symptoms, and then examine each hypothesis by searching in the direction of digraph arcs.

Expert Systems

This subsection discusses expert systems because of its relevance to FDD and by extension the current work. An expert system is used as a repository of knowledge and experience gained by process plant operators and control engineers. The fundamental laws of physics and chemistry can also be encoded for use in resolving diagnostic ambiguity and reaching a conclusion regarding the root- cause of a distributed, plant-wide disturbance.

Expert systems are computer-based applications used to deploy the knowledge, insights, advice and guidance of experts in a particular field. It implies that consistent expert knowledge can be formalized and reproduced when needed in the future for drawing conclusions. In building expert systems, the knowledge of domain expert is usually elicited through interviews with a knowledge engineer who later process and assemble the knowledge gathered in a knowledge base.

LINKman (ABB, 2002) is an expert system that has been successfully deployed in cement manufacturing plants and known to have brought about improvement in overall process performance, higher product quality and significant reduction in energy usage. A review of expert systems applications in a broad range of industries including mineral, chemical, nuclear power and brewing have been reported (Bearman and Milne, 1992).

Choices available for the development of expert systems include; (i) Conventional programming language (such as C++, C#, Java, C); (ii) AI programming language (especially, LISP and Prolog); and (iii) Expert system shells. Expert systems are usually deployed using expert system shells. Table 6 compares the various development tools available in order to make an informed decision about the suitability of one or more of the tools for a particular task.

Expert system tools have been developed and deployed for practical use in many engineering domain such as the process industry. Increase in the use of expert systems in various applications underscores the significant progress made since its inception as a unique branch of AI in the late 1960's from mere laboratory trials.

Based on comparisons in Table 6, it appears an expert system shell for encoding knowledge and physics of the process meets more of the diagnostic requirements in order to draw conclusions about the root-cause of process upset.

Table 6: Comparison of various development tools for expert system

	Conventional Programming Language (E.g. C++, C#, Java, C)	AI programming Language (especially, LISP & Prolog)	Expert System Shells
Development efforts/time required	High	Fair	Fair
Flexibility	High	High	Low
Inference power	Low	Low	High
Maintainability	Poor	Poor (usually requires input from specialists)	Very good
Explanation facility	Fair	Poor	Good
Graphics capability	poor	Poor	Good

A survey of some commercial expert system shells produces Table 7.

Table 7: Comparison of some commercial expert system shells

	G2®	JESS®	ACQUIRE®
Platform	Windows®, UNIX (Vs. 8.3 supports .NET)	Windows®, UNIX (Supports JVM not .NET directly)	Windows®, UNIX
Real time capability	Yes	No	No
License issues	Academic/ Educational use available: -20% of list price excluding maintenance - maintenance is 15% of list price -12 monthly use	Academic research only with no distribution rights	Academic/ Educational use with discount available
Embedability	Integrates with other applications	Usually with Java® applications	Integrates with other applications
Object oriented	Supports O-O technology	Supports O-O technology	No information
Inference mechanism	Forward and Backward chaining	No information	Forward and Backward chaining
Network capability	Excellent	Excellent-via Java® API	Needs further service subscription

2.9 World Experts and Centres of Excellence in Fault Diagnosis

Table 8 provides a list of experts in the field of process fault detection and diagnosis with their research centres. Also included in the table are particular topics of specialisation within the field.

Table 8: World experts and centres of excellence in process FDD

Institution	Research area(s)	Researcher
Centre for Process Systems Engineering Department of Chemical Engineering and Chemical Technology Imperial College South Kensington, London United Kingdom URL: http://www3.imperial.ac.uk	Process automation, Plant-wide disturbance detection and characterisation, Linear and non-linear root-cause diagnosis e.g. oscillations due to disturbance and limit cycles, Spectral decomposition and clustering e.g. PCA, ICA, NMF, auto covariance, surrogates, Control loop performance evaluation and assessment utilizing spectra analysis	Thornhill, N. F.
Department of Chemical & Biological Engineering Illinois Institute of Technology Chicago, Illinois, USA URL: http://www.iit.edu	Multivariable Process Monitoring, Bifurcation Analysis and Complexity in Large Distributed Systems, Performance Assessment, Fault Diagnosis, Fault Tolerant Control Multi-agent System for Modelling, Supervision and Control of Distributed Adaptive Systems	Cinar, A.
Purdue University School of Chemical Engineering Forney Hall of Chemical Engineering 480 Stadium Mall Drive West Lafayette, IN 47907-2100 URL: http://www.purdue.edu	Artificial Intelligence, Statistical Mechanics, Complex Adaptive Systems, Flow sheet Analysis, Fault Diagnosis, SDG	Venkatasubramanian, V.

Department of Chemical Engineering and Materials Science University of Southern California 925 Bloom Walk, HED 211 Los Angeles, CA 90089-1211 USA URL: http://chems.usc.edu	Process Monitoring, Fault Diagnosis, System Identification Data Analysis & Control, Model Predictive Control, Sensor Validation	Qin, S. J.
Department of Bioengineering & San Diego Supercomputer Center University of California, San Diego, 9500 Gilman Drive, La Jolla CA 92093-0412. URL: http://www.sdsc.edu/~mano/index.html Formerly at: Department of Chemical Engineering Purdue University Indiana, USA URL: http://www.purdue.edu	Flow sheet Analysis, SDG, Process Design, Control, Monitoring and Fault Diagnosis, Modelling, Simulation and Optimization, Data Mining and Statistical/Machine Learning, Artificial Intelligence, Parallel Computing	Maurya, M. R.
Department of Chemical Engineering MIT Cambridge, USA URL: http://www.mit.edu	Fault Diagnosis SDG (Active research in the 80s to mid 90s)	Kramer, M. A.
MIT 77 Massachusetts Ave, Rm. 66-372, Cambridge, MA 02139 http://web.mit.edu/braatzgroup/index.html Formerly at: Chemical & Biomolecular Engr. University of Illinois at Urbana-Champaign Urbana, Illinois USA URL: http://www.uiuc.edu	Fault detection using statistical analysis of process data simulation, design, and control of multiscale systems	Braatz, R.D.

Department of Chemical Engineering Texas Tech University 6th and Canton Mail Stop 3121 Lubbock, TX 79409-3121 URL: http://www.depts.ttu.edu/che/faculty/rrengasamy/rrengasamy.php Formerly at: Department of Chemical Engineering Clarkson University Potsdam, New York, USA URL: http://www.clarkson.edu	Multi-Scale Modelling and Optimization, Controller Performance Assessment and Process Fault Diagnosis , Flow sheet Analysis, SDG	Rengaswamy, R.
Chemical and Materials Engineering, University of Alberta Edmonton, Alberta CANADA T6G 2G6 http://www.ualberta.ca	Multivariate statistical analysis of plant data for control loop performance assessment, process monitoring, fault diagnosis, and development of new control-relevant identification algorithms for use in the design of model-based predictive controllers. Computer Process Control Process Automation	Shah, S. L.
University of Michigan Computer Science and Engineering 2260 Hayward Street Ann Arbor, MI 48109-2121 URL:http://eecs.umich.edu/~kuipers/ Formerly at: Department of Computer Science University of Texas at Austin Round Rock, Texas 78681 Austin, Texas http://www.cs.utexas.edu	Qualitative Modelling and Simulation	Kuipers, B.

Department of Chemical Engineering University of Texas Texas, USA URL: http://www.engr.utexas.edu	Control System Monitoring, Single and Multi Loop PID, MPC	Edgar, T. F.
School of Electrical and Electronic Engineering The University of Manchester Manchester M13 9PL http://www.eee.manchester.ac.uk	Multivariate statistical process control, Model predictive control, Control loop monitoring, Monitoring and control of batch processes, Flow assurance in the oil and gas industry.	Barry Lennox
Department of Chemical Engineering University of Texas at Austin Texas, USA URL: http://www.che.utexas.edu	Artificial Neural Networks for Fault Diagnosis and Data Rectification	Himmelblau, D. M.
Department of Chemical Engineering MIT Cambridge, USA URL: http://www.mit.edu	SDG, Extended SDG (ESDG) (80's and early 90's)	Oyeleye, O.O.
Department of Chemical Engineering Norwegian University of Science and Technology (NTNU) Norway URL: http://ww.ntnu.no/english	Plant-wide control and Optimization, Control structure design, distillation column design, Process control and dynamics	Sigurd Skogestad
School of Engineering University of Glasgow James Watt South Building, Glasgow G12 8QQ, Scotland United Kingdom http://www.gla.ac.uk/schools/engineering/staff/johnhowell/	Plant-wide fault detection and diagnosis and other abnormal situations in process plants-nuclear and chemical	John Howell

2.10 Chapter Summary

This chapter has explored existing approaches to process modelling namely quantitative model-based, qualitative model-based and process history-based methods. The chapter described graph theory as one of the fundamentals of the thesis. The chapter extended discussions on graphs to usability of directed graphs with respect to process schematics and in the storage and processing of graphs representations in computer programs. The relevance of connectivity matrix and XML were highlighted because of their central importance to the thesis.

Chapter 2 has also provided a review of existing work in the field of FDD. The chapter considered the application of a combination of approaches to develop FDD tools. One important observation that will find usage in the thesis is that a combination of techniques (hybrid), such as data-driven techniques using statistical methods and expert systems, is widely used and proved to be successful in developing FDD tools.

The chapter concluded with a table listing the centres of excellence, researchers and area of specialisation in FDD which provided a useful guide to where research activities are going on in the field across the globe.

3 Access to Engineering Information for Process Operations

This chapter describes the process of utilizing engineering information and data for process plant operational purposes. The chapter details all pertinent standards and commercial vendors of computer aided design (CAD) tools capable of producing intelligent piping and instrumentation diagrams (iP&IDs).

The chapter explores the concept of iP&ID as background to the research. International standards relevant to iP&IDs are ISO 10303-221 (also called AP221), ISO 15926 and IEC PAS 62424. The ISO 15926 is used in this thesis because this standard has been widely implemented by leading commercial CAD vendors. IEC PAS 62424 is a similar standard to ISO15926 and is discussed in details in this chapter as well. The reason for delving into IEC PAS 62424 is that it was extensively used at the early stages of the research. These standards permit interchange of iP&ID information between different tools. The chapter also describes the relevance of these standards to the thesis.

The issue of legacy CAD drawings, such as those drawn with AutoCAD®, which may require conversion to intelligent P&IDs is also discussed. Finally, the chapter considers the requirements of the project and reviews computer-aided tools for the preparation and export of iP&IDs in a form suitable for use within the project. A decision is needed about whether to employ such a tool and which one to choose.

An illustrative example is included to demonstrate the capability of the tool developed in accessing and using intelligent P&ID information.

3.1 Engineering Data and Information during Process Design and Construction

A vast amount of engineering design data and information are generated during the construction of a process plant. The data and information are contained in documents which become part of important industrial intellectual property (Daratech, 2004). An example of such document is a process P&ID. After the plant has been commissioned and is on-stream, most of these design data and information are seldom put into use mainly due to the storage and retrieval issues.

3.1.1 Motivation and Need for Interoperability

In 2002, the National Institute of Standards and Technology (NIST, 2004) estimated the annual cost of poor interoperability and data exchange in the US capital facilities industry alone at USD 15.8 billion while the McGraw Hill ENR Technology for Construction 2007 report on interoperability estimates the cost to be twice as much as the 2002 NIST report (ENR, 2005). Standards have been and are being developed to deal with the problem of interoperability and ensure more efficient systems integration and workflow.

With the advancement in computer processing power capability, decreasing digital media storage device coupled with rapid evolution of information technology, process industries and CAD tool vendors are devising means to leverage the engineering information for better information management. These developments include maintaining data integrity and management of change, enhancing information exchange, and promoting work flow in multi-disciplinary project.

The advancement in process information technology has led to development of what is now referred to as intelligent P&IDs which are traditional drawings with data repositories that are not visible on the graphic coupled with relevant associated international standards. Engineering data and information that were, hitherto, inaccessible for operational purposes are now possible such as the use of XML to store, retrieve, transmit and exchange engineering data and information.

3.2 Intelligent P&IDs

A P&ID is a drawing of a process showing items of equipment and the connections between them. It generally conforms to established conventions and standards for the layout and symbols. An *intelligent* P&ID, example shown in Figure 37, stores additional information and is able to exchange information about the items, their layout and connections with a data base and with other engineering tools.

The ability of iP&ID to exchange drawing data and information in a text-based format as XML allows computer processing and algorithmic manipulation for automating the generation of process connectivity descriptions such as digraphs and connectivity matrices which are needed for the project.

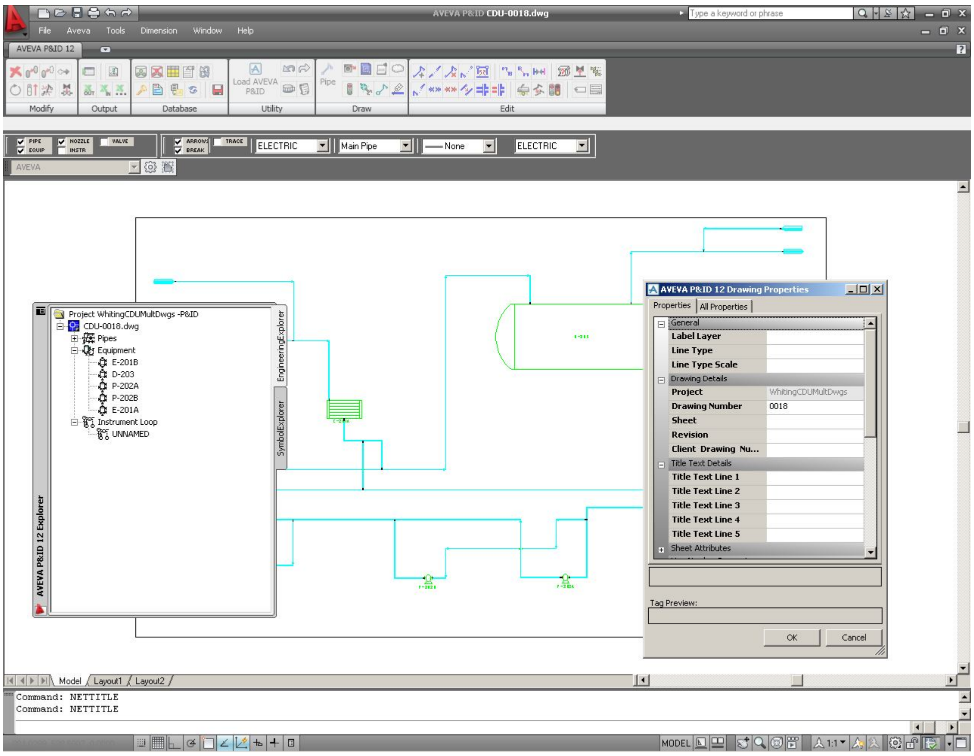


Figure 37: An intelligent P&ID drawn with AVEVA P&ID® CAD tool running on AutoCAD®

The core concept behind an intelligent P&ID is the *data model*. The data model is a structured description of types of objects and the properties of specific objects. An iP&ID stores tagged items, quantities, connectivity and directionality data independently of the drawing and hence can export information about items of equipment and the links between them into other formats. Most iP&ID tools store their data model in a proprietary format, but there is user-led pressure for future tools to comply with open standards which most leading commercial CAD vendors are subscribing to trend (Laud, 2011).

The main commercial products for creating iP&IDs are:

- AutoCAD P&ID which is based on AutoCAD
- Aveva P&ID which is based on AutoCAD
- Bentley AutoPlant P&ID which is based on AutoCAD
- Innotec Comos P&ID
- Intergraph SmartPlant

Some features of these commercial CAD tools will be reviewed in Section 3.7

3.3 Options for Converting *dumb* P&IDs to *intelligent* P&IDs

Traditional drawings such as legacy drawings in earlier versions of AutoCAD tools are called *dumb* P&IDs and they do not store additional information besides the arcs and line graphics shown in the drawings. Old and extant P&IDs (or control and flow diagrams, CFDs) lack the data model which makes a CAD intelligent. With most brown field process plants designed and drawn using traditional CAD drawing tools, there is, obviously an issue of the conversion of the legacy dumb P&IDs to intelligent P&IDs. Effective means of dealing with and managing legacy data issues are unarguably the biggest problem facing the process industries (Gartner, 2008). Options available for converting *dumb* P&IDs to *intelligent* P&IDs include manual, semi-automatic and automatic conversions. These options will be explored in the succeeding sections.

3.3.1 Manual Conversion

The amount of work depends on how old the drawing is and whether the engineer conformed with good practices when creating the drawing. All the tools listed in section 3.7 offer some assistance with conversion. In general, they use rules to identify items of equipment and can partially populate a database. The user may have to intervene and make some manual decisions. The amount of effort required is greatest here and the possibility for the draughtsman to make a mistake is higher. Further, extant drawings might have an outdated standard of practice.

3.3.2 Automated Conversion

Several CAD vendors claim to have tools for automatic conversion of dumb P&IDs to full or partial intelligent P&IDs on their respective websites. For example, the Noumenon tools can partially convert dumb AutoCAD® drawings (Laud, 2011). At present, a data base can be populated with items of equipment but direction links are not extracted. These tools are under development, however. Intergraph offers a commercial conversion service from dumb AutoCAD® drawings to SmartPlant® drawings for a fee plus a one-off set-up fee. The service is partly automated and the manual steps are done by Intergraph.

3.4 Relevant Standards

There are standards relevant to iP&IDs, and which also cover the concept of an engineering data warehouse (EDW). Computer-aided tools which combine iP&ID and EDW offer the generation and open interchange of information about items of process equipment and their connections.

3.4.1 ISO 15926

POSC-Caesar Association (PCA, 2011), a global, non profit organization initiated ISO15926 and is committed to its development, maintenance & enhancement. ISO 15926 is an international organisation for standardization (ISO) standard for seamless exchange and integration of industrial data. ISO 15926 is formally called *Industrial automation systems and integration - Integration of life-cycle data for process plants including oil and gas production facilities*. ISO 15926 is implemented using standards from worldwide consortium (W3C) and standardizes terminology, information organization and how systems connect and exchange information.

The development project within ISO is undertaken and overseen mainly by the advancing development of ISO15926 (ADI) and intelligent data sets (IDS) groups. It specifies a data model for information for the engineering, construction and operation of process plants. The aim is to mitigate the current high cost of rekeying and reformatting information to move it from one proprietary system to another (PCA, 2011). This data model is the engineering data warehouse defined earlier. A key feature illustrated in Figure 38 is the sharing and integration of information amongst all parties involved in the plant's life cycle including contractors, automation vendor and end-user.

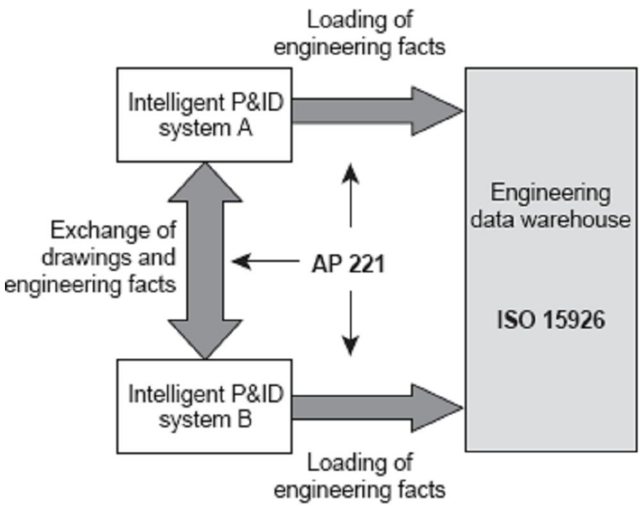


Figure 38: Relationship between intelligent P&IDs and an engineering data warehouse, cited from Leal, 2005

Information that enables engineering projects to proceed smoothly such as parts lists and process diagrams (intelligent P&IDs) are built from information in the data model. For instance, querying an item such as a pump in a drawing or parts list should bring up additional information about it from the data base, for instance the type and size of pump and the manufacturer. The data model can be populated from a drawing and can be updated by making changes to the drawings.

3.4.2 ISO 10303-221

ISO10303-221 *Functional Data and Their Schematic Representation for Process Plants* (also called AP221) is a protocol for the exchange of intelligent schematics between different software applications. It can be used for the exchange of iP&IDs between a contractor and a plant owner or between an automation supplier and the contractor. ISO10303-221 specifies the construction of drawings such as P&IDs and also is used for exchange of data between an intelligent P&ID system and an EDW defined in accordance with ISO15926. Figure 39 shows the way in which ISO1030 and ISO 15926 overlap to achieve this task. As indicated in Figure 39, the drawing captures the plant as it is, there is a shared data model containing information about the items of equipment in the drawing, and ISO 15926 records both the present and past states of the plant. According to Leal (2005) from whom the illustrations have been taken, the ability of the ISO 15926 data model to record changes is its defining feature.

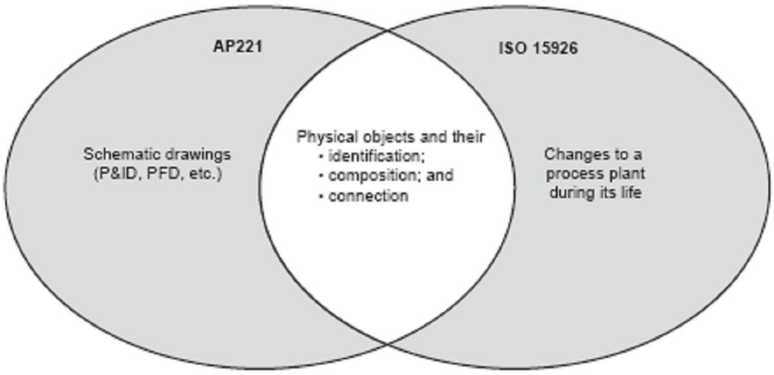


Figure 39: The roles of AP221 (ISO10303-221) and ISO 15926 in maintaining information about the plant, taken from Leal, 2005

3.4.4 IEC PAS 62424

IEC PAS 62424 is called *Representation of process control engineering requests in P&IDs and data exchange between P&ID tools and PCE - CAE (Process Control Engineering - Computer Aided Engineering tools).* IEC/PAS stands for international electrotechnical commission/publicly available specification. This specification has a focus on control and instrumentation and describes how process control engineering functionality is to be represented in P&IDs. The idea is to indicate the category and processing function, independent from the physical realization (distributed control systems (DCS), wireless, pneumatic). It also specifies the flow of data between a P&ID tool and a process control engineering (PCE) tool by means of an XML data transfer language called computer aided engineering exchange (CAEX). CAEX specifies items of equipment and directional links between them. This standard is discussed further with some examples in section 3.10.

3.4.5 Other Standards

There are also other relevant standards that specify the symbols to be used in P&IDs and naming/numbering conventions. These standards are:

- ISO 10628 Graphical symbols for process equipment
- ISO 14617 Graphical symbols for diagrams

3.5 History of the Standards

This subsection provides an overview of the standardization. It provides the evolutionary nature of the standards as well as the efforts put into their development at various stages by working groups responsible for specifying the standards.

3.5.1 ISO 10303 and ISO 15926

The ISO 10303 and 15926 standards originated with POSC Caesar, a non-profit European organization that works on open specifications for standards to allow interoperability of data and software. A US industry consortium called FIATECH has adopted ISO 15926 as a way to integrate and automate the execution of large capital projects. Since 2010 POSC Caesar has been collaborating with FIATECH to implement a joint operational reference data (JORD) as a more robust alternative to the existing POSC Caesar openly available reference data library (RDL).Both POSC Caesar and FIATECH are developing case studies and demonstrations.

3.5.2 IEC PAS 62424 and CAEX

IEC 62424 and CAEX emerged from an initiative called AutomationML®. The objectives are similar to those above, namely to establish an open format to allow communication between different engineering tools. The emphasis has been on mechanical, electrical, robotics, and automation equipment. CAEX describes the properties and relations of objects in their hierarchical structure and is the data transfer language for carrying information between different engineering tools.

3.5.3 Harmonization between CAEX and ISO 15926

The ISO 15926 and CAEX approaches seem to be developing in parallel and without reference to each other and harmonization between ISO 15926 and CAEX is needed (Koning, 2007). Similarly, a Google® search for pages including both AutomationML® for CAEX and XMpLant for IS0 15926 implementations returned no hits as the time of writing this thesis.

3.6 Data Export and Interchange

This section describes data transfer and information exchange formats among various engineering design tools as executed in compliance with ISO 15926 and IEC PAS 62424 and CAEX.

3.6.1 Data Exchange in ISO 15926

The specifications in Part 7 of ISO 15926 standard deals with methods of data transfer and information exchange between different engineering tools. It is based on languages used for the world-wide web called resource description framework (RDF) and web ontology language (OWL). ISO 15926 Part 7 has some way to go before full implementation, however.

3.6.2 Data Exchange using CAEX

CAEX is an XML schema for export from tools that comply with the IEC PAS 62424 standard(Fedai and Drath, 2005). As far as can be ascertained, only Comos P&ID is known to have XML export that conforms to the CAEX schema.

3.6.3 XMpLant Export

An organization called Noumenon Consulting Inc. provides ISO 15926-compliant data transfer tools based on XML according to a schema called XMpLant. Noumenon provides interchange of intelligent P&IDs between several of the iP&ID tools listed in section 3.2. Figure 40, from the web pages of Noumenon Consulting Inc. shows how mapping files are used to convert between the native formats and the open XMpLant format.

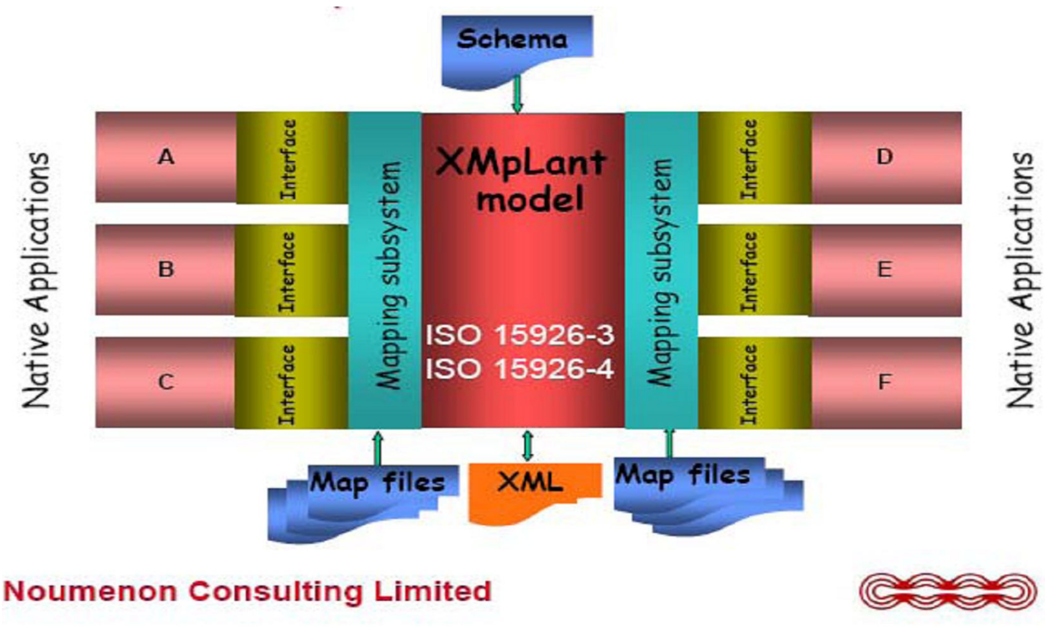


Figure 40: XMpLant data model overview showing the mapping of intelligent P&IDs between the P&ID applications of different vendors. Cited from Laud (2006)

3.6.4 Data Exchange by Other Means

Most of the P&ID tools listed in Section 3.7 have proprietary formats for storing and representing their data model and for export of the data model (i.e. of the equipment lists and connections). The web site of AutoCAD P&ID, for instance, says that the model is exported in an Excel spreadsheet.

3.7 A Review of Commercial Products

Table 9 outlines commercial products that meet some or all of the requirements of the project, namely:

- to handle intelligent P&IDs,
- to provide export of the items of equipment and links between them,
- to convert dumb P&IDs to intelligent P&IDs.

Table 9: Comparison of commercial products against their level of suitability for the requirements of the research project

Tool	Intelligent P&ID	Standards	Export	Dumb-Intelligent Conversion
SmartPlant P&ID	Own intelligent P&ID and Oracle or MSSQL data base	Fully ISO 15926 compliant , see http://www.intergraph.com/ppm/iso15926.aspx	Noumenon XMpLant is available	Service available from Intergraph at a fee
Comos P&ID	Own Intelligent P&ID and Oracle or MSAccess or MSSQL data base	Moving to ISO 15926, working with FIATECH	XML (CAEX schema)	Users can convert dumb AutoCAD P&ID
Bentley AutoPlant P&ID	Based on AutoCAD 2008 with Intelligent P&ID features	Not ISO 15926, proprietary system	Noumenon XMPlant export is not yet available	Users can convert dumb AutoCAD drawing

Bentley OpenPlant P&ID	Own intelligent P&ID and a data base. It is not available until 2009	Fully ISO 15926	Export to any ISO 15926 data warehouse	None, but it can import from AutoPlant
Aveva VPE P&ID	Based on AutoCAD 2010 with Intelligent P&ID	Not ISO15926	ISO 15926 and XMpLant are available	Users can convert dumb AutoCAD
AutoCAD P&ID	Based on AutoCAD 2010, Intelligent P&ID	Not ISO 15926	Noumenon XMPlant export is available	Users can convert dumb AutoCAD
XMplant	It is a mapping tool that converts proprietary formats to and from ISO 15926.	Fully ISO 15926	ISO 15926 XMPlant Schema	Can partially convert dumb AutoCad (items only, no connections). It can read intelligent P&IDs from Aveva and AutoCAD P&ID

3.8 Discussion of CAD Commercial Tools within the Context of the Project Requirements

The project has a basic requirement to extract a list of *plant items and* the *directional connections* between them from a CAD drawing. The research has indicated that besides a full process P&ID, a high-level drawing of the process such as a process flow diagram (PFD) or control and flow diagram (CFD) can also be used to obtain connectivity and directionality information from process topology. Items to be extracted include equipment, signal lines, instruments and controllers as well as pipes, and piping components.

At the Academic Level: The price of commercial tool and interest in academic collaboration is a major determinant in choosing the appropriate software tool for the research.

At the Industrial Level: The selected method should be automated, future-proof and also be able to handle legacy drawings.

3.8.1 Software Selection

At the Academic Level: The research work has demonstrated a range of solutions for getting a connectivity description from a P&ID. These include CAEX based on IEC PAS 62424 standard and XMpLant schemas based on ISO15926 standard.

At Industrial Level: There would be the need to consider long-term solution for getting connectivity information from CAD drawings. The export from selected CAD tool should conform to a widely used international standard and should be vendor and platform independent.

For Both: Given the requirement to re-use legacy AutoCAD drawings, one of the intelligent P&ID tools based on AutoCAD appears to be a good choice.

The project sponsor prefer ISO15926 standard to CAEX 62424 because ISO15926 is the main standard used in North America and has been implemented by more major CAD tool vendors than the CAEX standard. The wider coverage and implementation of ISO 15926 provided a much more challenging PhD project.

Aveva and Noumenon Consulting are both suitable project partners to deploy intelligent CAD tool. These considerations informed our decision to choose Aveva® P&ID software tool for creating intelligent P&ID in this research.

3.9 Generation of a Connectivity Description of a Process

For the purposes of the research project, any export of an intelligent P&ID that contains the information about equipment and instrumentation with directional links will suffice. Options for generating process connectivity from plant structure include:

- Using an AutoCAD P&ID Excel description of the drawing
- Parsing an XML description of the drawing (CAEX or XMpLant schema)
- Creation of the connectivity information by hand

The XML description of the process is chosen because XML is becoming the *de facto* neutral standard for information storage, retrieval and transmission in the industry. For example, the adoption of Microsoft Office open XML formats such as docx, xlsx and pptx extensions is an eloquent testimony of the acceptance of XML standards(Microsoft, 2006).

3.9.1 ISO 15926 XML Output

This section presents a simple intelligent process P&ID with a level control loop. The purpose for its inclusion is to show how various elements of process schematics are described and represented using text based XML compliant with ISO15926 and XMpLant Schema(Laud, 2011) implemented in AVEVA P&ID®.

The section discusses connectivity representation as well as description of individual components shown visually in Figure 41. A control loop is described as a composition of the various elements making up the loop, namely, measurement device, controller and final control element such as a pneumatic valve.

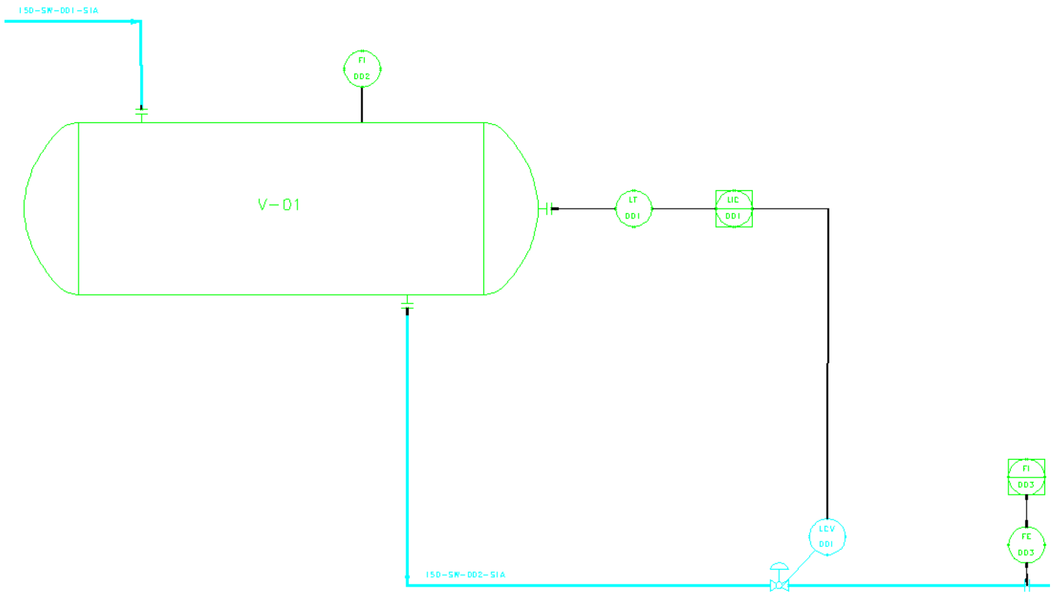


Figure 41: Process schematic created with AVEVA P&ID®

Connectivity Description in XML

In Figure 41, temperature indicator *TI-002(35CB)* is connected to the vessel *V-01(346F)* by a process link *(25D6)* with no nozzle. This shows up in the XML as a virtual nozzle (*346F-35D6). System generated unique identifiers are quoted in brackets. The process link *(35D6)* shows the connection between the vessel *V-01 (346F)* and *TI-002 (35CB)* XML description of the connectivity described above is given below.

Below are the details of the loop L-001 and constituent instruments.

The L-001 loop XML description:


```xml
- <InstrumentLoop ID="XMP_8" TagName="L-001">
  - <GenericAttributes Set="" Number="13">
      <GenericAttribute Name="Block" Format="string" Value="" />
      <GenericAttribute Name="TagBlock" Format="string" Value="" />
      <GenericAttribute Name="Function" Format="string" Value="" />
      <GenericAttribute Name="TagFunction" Format="string" Value="" />
      <GenericAttribute Name="Prefix" Format="string" Value="" />
      <GenericAttribute Name="TagPrefix" Format="string" Value="" />
      <GenericAttribute Name="Type" Format="string" Value="L" />
      <GenericAttribute Name="TagType" Format="string" Value="L" />
      <GenericAttribute Name="Number" Format="string" Value="001" />
      <GenericAttribute Name="TagSequenceNo" Format="string" Value="001" />
      <GenericAttribute Name="Suffix" Format="string" Value="" />
      <GenericAttribute Name="TagSuffix" Format="string" Value="" />
      <GenericAttribute Name="Grid_Reference" Format="string" Value="" />
    </GenericAttributes>
    <Association Type="is a collection including" ItemID="XMP_36" Name="ProcessInstrument" Context="ISOExp3" />
    <Association Type="is a collection including" ItemID="XMP_29" Name="ProcessInstrument" Context="ISOExp3" />
    <Association Type="is a collection including" ItemID="XMP_18" Name="ProcessInstrument" Context="ISOExp3" />
  </InstrumentLoop>
```


Loop L-001 constituents

Loop L-001 constituents

LT-001 XML description


```xml
- <ProcessInstrument ID="XMP_18" TagName="LT-001" ComponentClass="LevelTransmitter" StockNumber="IIN" Specification="">
  - <Extent>
      <Min X="439.65" Y="583.65" Z="0.00" />
      <Max X="456.35" Y="600.35" Z="0.00" />
    </Extent>
    <Presentation Layer="AS_INST" LineType="Solid" LineWeight="0.25" Color="Lime" R="0" G="1" B="0" />
  - <Position>
      <Location X="448.00" Y="592.00" Z="0.00" />
      <Axis X="0.00" Y="0.00" Z="1.00" />
      <Reference X="1.00" Y="0.00" Z="0.00" />
    </Position>
    <PersistentID Identifier="3486" Context="ISOExp3" />
```


LIC-001 XML description


```xml
- <ProcessInstrument ID="XMP_29" TagName="LIC-001" ComponentClass="LevelIndicatorController" StockNumber="SDC1" Specification="">
  - <Extent>
      <Min X="479.65" Y="583.65" Z="0.00" />
      <Max X="496.35" Y="600.35" Z="0.00" />
    </Extent>
    <Presentation Layer="AS_INST" LineType="Solid" LineWeight="0.25" Color="Lime" R="0" G="1" B="0" />
  - <Position>
      <Location X="488.00" Y="592.00" Z="0.00" />
      <Axis X="0.00" Y="0.00" Z="1.00" />
      <Reference X="1.00" Y="0.00" Z="0.00" />
    </Position>
    <PersistentID Identifier="34CA" Context="ISOExp3" />
```

LCV-001 XML description


```
- <ProcessInstrument ID="XMP_36" TagName="LCV-001" ComponentClass="LevelControlValve" StockNumber="DABAVA" Specification="S1A">
  - <Extent>
      <Min X="523.65" Y="496.50" Z="0.00" />
      <Max X="532.35" Y="509.00" Z="0.00" />
    </Extent>
    <Presentation Layer="AS_PIPE" LineType="Solid" LineWeight="0.25" Color="Aqua" R="0" G="1" B="1" />
  - <Position>
      <Location X="528.00" Y="499.00" Z="0.00" />
      <Axis X="0.00" Y="0.00" Z="1.00" />
      <Reference X="1.00" Y="0.00" Z="0.00" />
    </Position>
    <PersistentID Identifier="3527" Context="ISOExp3" />
```


Connections from V-01 to TI--002


```
- <PipingNetworkSegment ID="XMP_85" TagName="" ComponentClass="PipeLine" Specification="">
  - <Extent>
      <Min X="368.85" Y="630.00" Z="0.00" />
      <Max X="369.15" Y="643.00" Z="0.00" />
    </Extent>
    <Presentation Layer="AS_PIPE" LineType="Solid" LineWeight="0.25" Color="White" R="1" G="1" B="1" />
    <PersistentID Identifier="35D6" Context="ISOExp3" />
  - <GenericAttributes Set="PipeLine" Number="24">
      <GenericAttribute Name="BranchId" Format="string" Value="90" />
      <GenericAttribute Name="Source" Format="string" Value="346F-35D6" />
      <GenericAttribute Name="Destination" Format="string" Value="35CB" />
      <GenericAttribute Name="Connected_PDMS_Type_From" Format="string" Value="" />
      <GenericAttribute Name="Connected_PDMS_Type_To" Format="string" Value="INST" />
      <GenericAttribute Name="Connected_Handle_From" Format="string" Value="346F-35D6" />
      <GenericAttribute Name="Connected_Handle_To" Format="string" Value="35CB" />
      <GenericAttribute Name="OwnerId" Format="string" Value="35E6" />
      <GenericAttribute Name="Size" Format="string" Value="" />
      <GenericAttribute Name="Fluid" Format="string" Value="" />
      <GenericAttribute Name="Number" Format="string" Value="" />
      <GenericAttribute Name="Specification" Format="string" Value="" />
      <GenericAttribute Name="Insulation_Table" Format="string" Value="" />
      <GenericAttribute Name="Tracing_No" Format="string" Value="" />
      <GenericAttribute Name="Insulation_Condition" Format="string" Value="" />
      <GenericAttribute Name="Paint_Code" Format="string" Value="" />
      <GenericAttribute Name="Project1" Format="string" Value="" />
      <GenericAttribute Name="Project2" Format="string" Value="" />
      <GenericAttribute Name="Project3" Format="string" Value="" />
      <GenericAttribute Name="Area" Format="string" Value="" />
      <GenericAttribute Name="Insulation_Index" Format="string" Value="" />
      <GenericAttribute Name="Tracing_Size" Format="string" Value="" />
      <GenericAttribute Name="Tracing_Type" Format="string" Value="" />
      <GenericAttribute Name="REMARKS" Format="string" Value="" />
    </GenericAttributes>
    <Connection FromID="346F-35D6" ToID="35CB" ToNode="2" />
```


Virtual nozzle (346F-35D6). Not shown on the drawing


```
- <Nozzle ID="XMP_93" TagName="" ComponentClass="VirtualNozzle">
  - <Extent>
      <Min X="369.00" Y="643.00" Z="0.00" />
      <Max X="369.00" Y="643.00" Z="0.00" />
    </Extent>
    <Presentation Layer="DummyLayer" R="0" G="0" B="0" />
  - <Position>
      <Location X="369.00" Y="643.00" Z="0.00" />
      <Axis X="0.00" Y="0.00" Z="0.00" />
      <Reference X="1.00" Y="0.00" Z="0.00" />
    </Position>
    <PersistentID Identifier="346F-35D6" Context="ISOExp3" />
  - <Line>
      <Presentation Layer="DummyLayer" R="0" G="0" B="0" />
    - <Extent>
        <Min X="369.00" Y="643.00" Z="0.00" />
        <Max X="369.00" Y="643.00" Z="0.00" />
      </Extent>
      <Coordinate X="369.00" Y="643.00" Z="0.00" />
      <Coordinate X="369.00" Y="643.00" Z="0.00" />
    </Line>
    <NominalDiameter Value="0.00" Units="mm" />
  </Nozzle>
</Equipment>
```

Connections from V-01 to LIC-001

Nozzle to LT-001


```xml
- <PipingNetworkSegment ID="XMP_91" TagName="" ComponentClass="PipeLine" Specification="">
  - <Extent>
      <Min X="425.00" Y="591.85" Z="0.00" />
      <Max X="440.00" Y="592.15" Z="0.00" />
    </Extent>
    <Presentation Layer="AS_PIPE" LineType="Solid" LineWeight="0.25" Color="White" R="1" G="1" B="1" />
    <PersistentID Identifier="348F" Context="ISOExp3" />
  - <GenericAttributes Set="PipeLine" Number="24">
      <GenericAttribute Name="BranchId" Format="string" Value="10" />
      <GenericAttribute Name="Source" Format="string" Value="347C" />
      <GenericAttribute Name="Destination" Format="string" Value="3486" />
      <GenericAttribute Name="Connected_PDMS_Type_From" Format="string" Value="NOZZ" />
      <GenericAttribute Name="Connected_PDMS_Type_To" Format="string" Value="INST" />
      <GenericAttribute Name="Connected_Handle_From" Format="string" Value="347C" />
      <GenericAttribute Name="Connected_Handle_To" Format="string" Value="3486" />
      <GenericAttribute Name="OwnerId" Format="string" Value="34A9" />
      <GenericAttribute Name="Size" Format="string" Value="" />
      <GenericAttribute Name="Fluid" Format="string" Value="" />
      <GenericAttribute Name="Number" Format="string" Value="" />
      <GenericAttribute Name="Specification" Format="string" Value="" />
      <GenericAttribute Name="Insulation_Table" Format="string" Value="" />
      <GenericAttribute Name="Tracing_No" Format="string" Value="" />
      <GenericAttribute Name="Insulation_Condition" Format="string" Value="" />
      <GenericAttribute Name="Paint_Code" Format="string" Value="" />
      <GenericAttribute Name="Project1" Format="string" Value="" />
      <GenericAttribute Name="Project2" Format="string" Value="" />
      <GenericAttribute Name="Project3" Format="string" Value="" />
      <GenericAttribute Name="Area" Format="string" Value="" />
      <GenericAttribute Name="Insulation_Index" Format="string" Value="" />
      <GenericAttribute Name="Tracing_Size" Format="string" Value="" />
      <GenericAttribute Name="Tracing_Type" Format="string" Value="" />
      <GenericAttribute Name="REMARKS" Format="string" Value="" />
    </GenericAttributes>
    <Connection FromID="347C" ToID="3486" ToNode="2" />
```


LT-001 to LIC-001


```xml
- <PipingNetworkSegment ID="XMP_77" TagName="" ComponentClass="PipeLine" Specification="">
  - <Extent>
      <Min X="456.00" Y="591.85" Z="0.00" />
      <Max X="480.00" Y="592.15" Z="0.00" />
    </Extent>
    <Presentation Layer="AS_PIPE" LineType="Solid" LineWeight="0.25" Color="White" R="1" G="1" B="1" />
    <PersistentID Identifier="3652" Context="ISOExp3" />
  - <GenericAttributes Set="PipeLine" Number="24">
      <GenericAttribute Name="BranchId" Format="string" Value="110" />
      <GenericAttribute Name="Source" Format="string" Value="3486" />
      <GenericAttribute Name="Destination" Format="string" Value="34CA" />
      <GenericAttribute Name="Connected_PDMS_Type_From" Format="string" Value="INST" />
      <GenericAttribute Name="Connected_PDMS_Type_To" Format="string" Value="INST" />
      <GenericAttribute Name="Connected_Handle_From" Format="string" Value="3486" />
      <GenericAttribute Name="Connected_Handle_To" Format="string" Value="34CA" />
      <GenericAttribute Name="OwnerId" Format="string" Value="3662" />
      <GenericAttribute Name="Size" Format="string" Value="" />
      <GenericAttribute Name="Fluid" Format="string" Value="" />
      <GenericAttribute Name="Number" Format="string" Value="" />
      <GenericAttribute Name="Specification" Format="string" Value="" />
      <GenericAttribute Name="Insulation_Table" Format="string" Value="" />
      <GenericAttribute Name="Tracing_No" Format="string" Value="" />
      <GenericAttribute Name="Insulation_Condition" Format="string" Value="" />
      <GenericAttribute Name="Paint_Code" Format="string" Value="" />
      <GenericAttribute Name="Project1" Format="string" Value="" />
      <GenericAttribute Name="Project2" Format="string" Value="" />
      <GenericAttribute Name="Project3" Format="string" Value="" />
      <GenericAttribute Name="Area" Format="string" Value="" />
      <GenericAttribute Name="Insulation_Index" Format="string" Value="" />
      <GenericAttribute Name="Tracing_Size" Format="string" Value="" />
      <GenericAttribute Name="Tracing_Type" Format="string" Value="" />
      <GenericAttribute Name="REMARKS" Format="string" Value="" />
    </GenericAttributes>
    <Connection FromID="3486" ToID="34CA" ToNode="2" FromNode="2" />
```


High level Hierarchical Structure

The AVEVA® P&ID hierarchical description breaks down process plant drawing into a *PlantModel root* element. Under the root element are other sub-elements, along with their sub-elements, attributes and directional connections into the following pertinent elements:

- Piping NetworkSystem
- Equipments
- Instruments

Figure 42, Figure 44, Figure 45, Figure 46 and Figure 47 produced with Stylus Studio XML viewer show the high level hierarchical structure of process P&ID as described by AVEVA® P&ID ISO15926 XML export. The relevant top-level hierarchical topologies as defined by AVEVA® P&ID CAD tool are shown in the Figure 42 to Figure 47 .

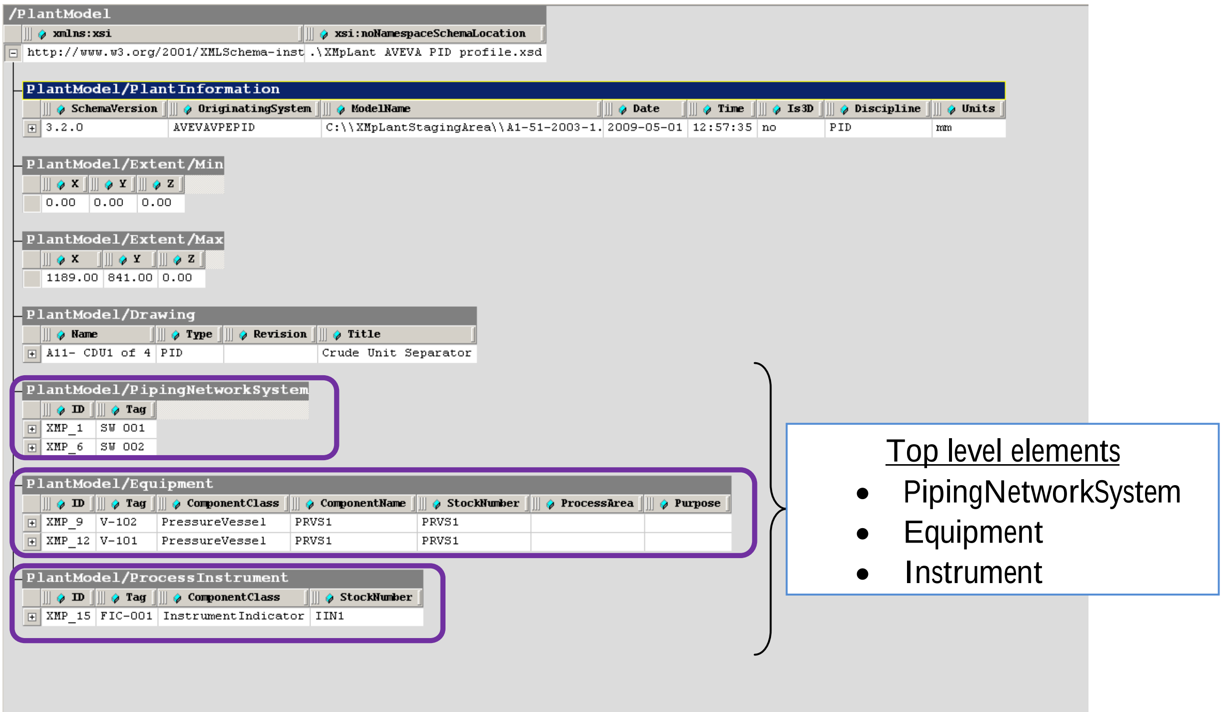


Figure 42 : Overall plant hierarchical structure in Stylus Studio XML viewer

The reason for showing the hierarchical structure is that it gives a clearer picture of the XML structure in a condensed form when compared with the XML text in Figure 43

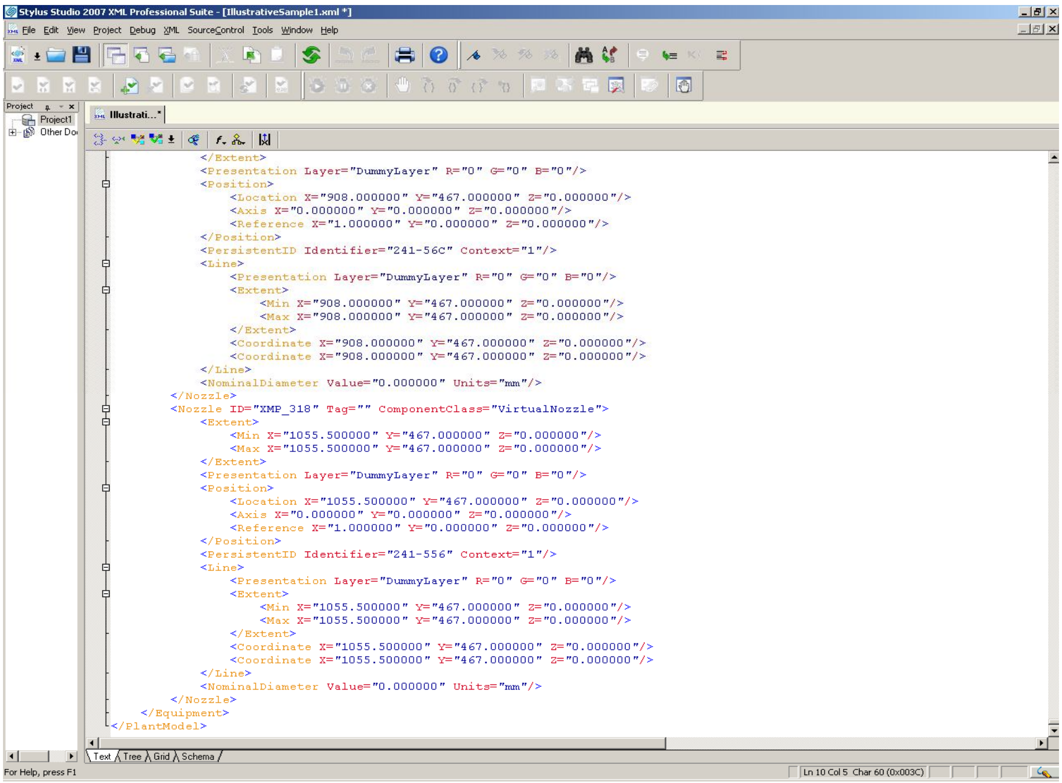


Figure 43: An example of XML text output

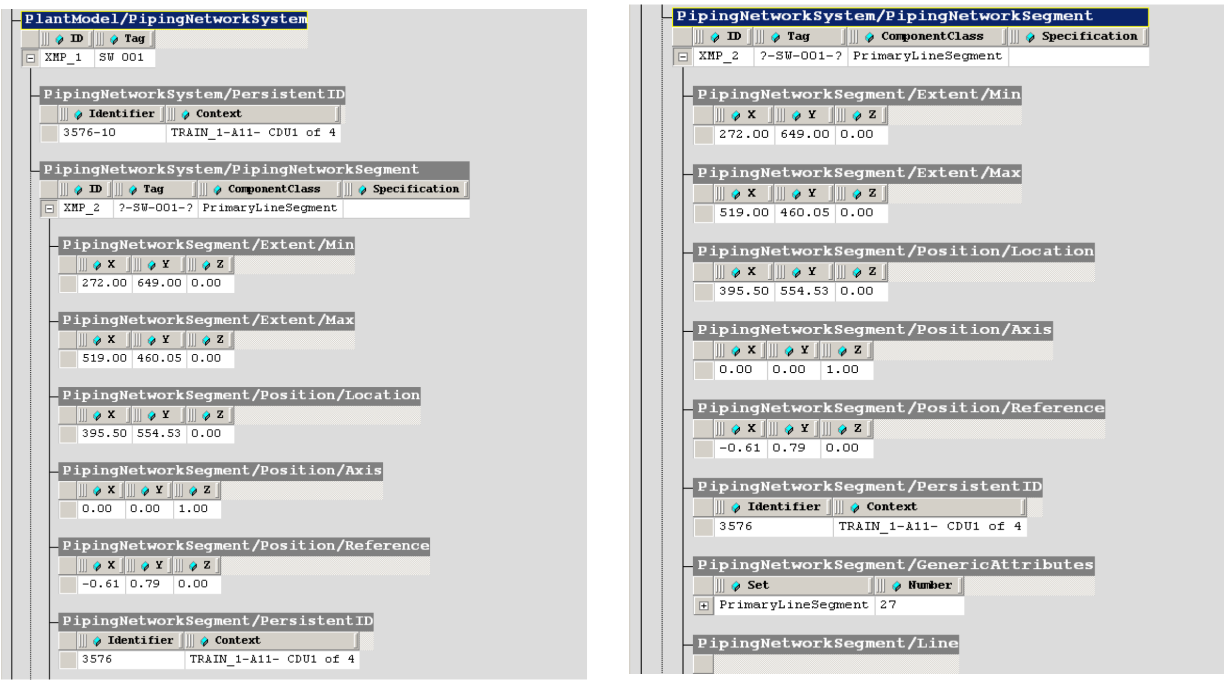


Figure 44: PipingNetworkSystem (left panel) and PipingNetworkSegment (right panel). PipingNetworkSegment is an element under PipingNetworkSystem

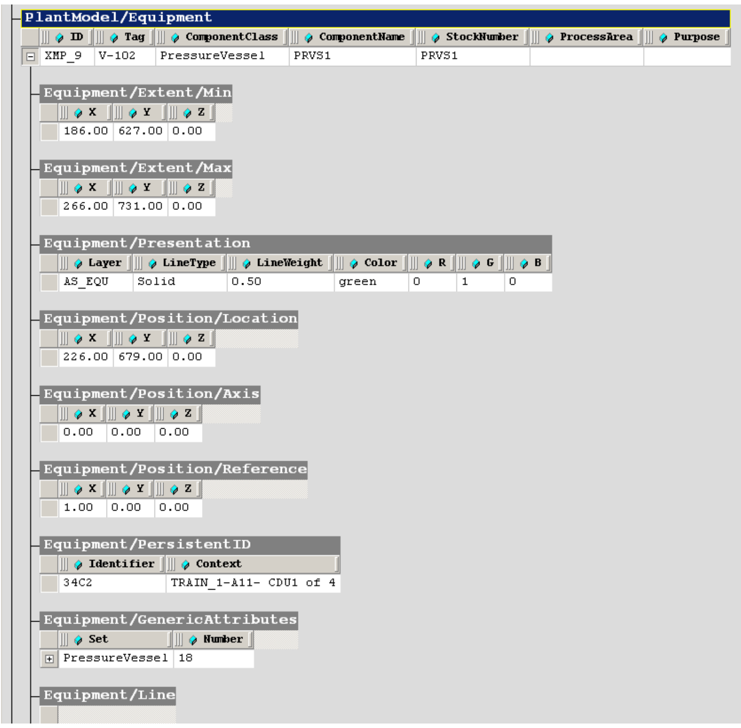


Figure 45: Process equipment and attributes

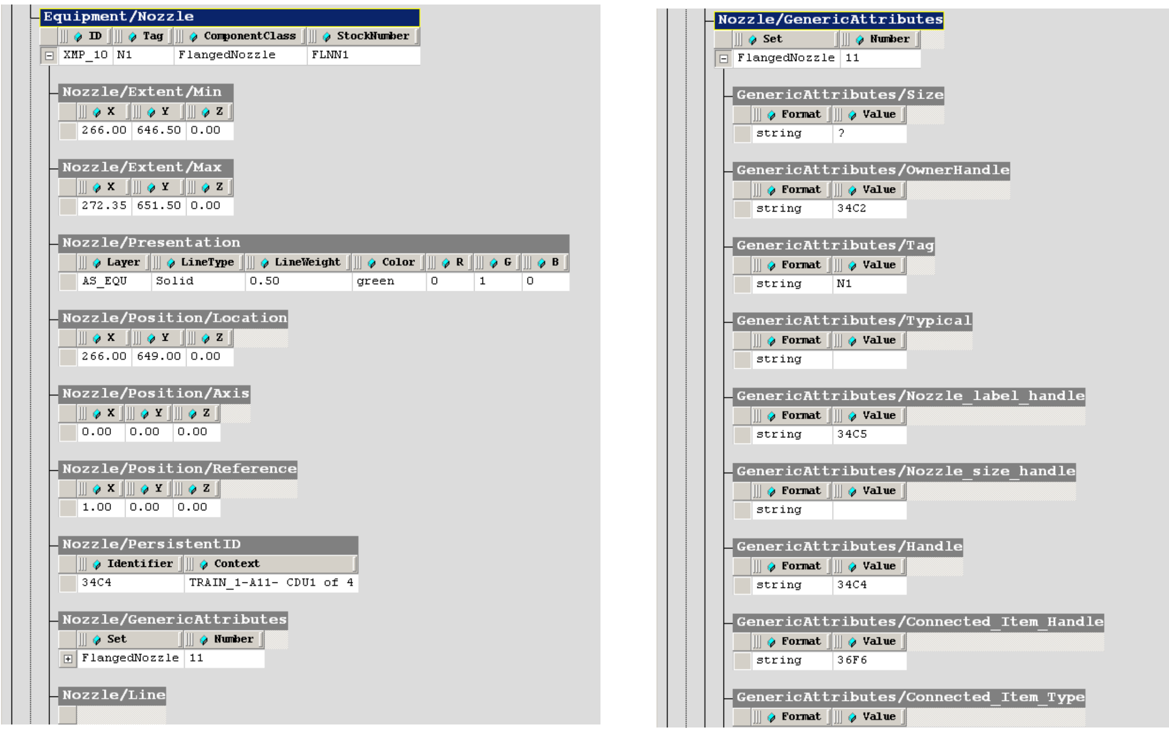


Figure 46: Nozzle (left panel) and nozzle attributes

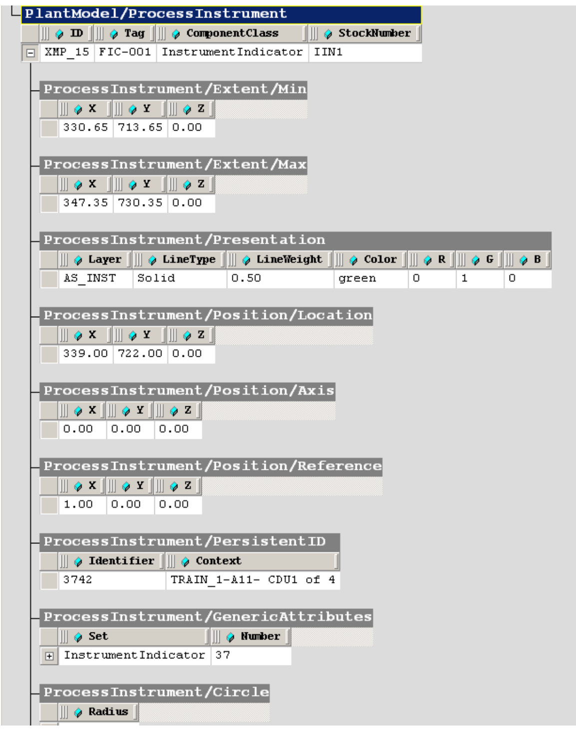


Figure 47: Instrument item and attributes

3.9.2 Connectivity Matrix from XML

The XML file produced from intelligent CAD drawing contains all the necessary information required to create a connectivity matrix. An example is shown in Figure 48. The process of extracting relevant data from XML file is known as *parsing.* The process of creating a connectivity matrix from an XML description of process plant with "*Process Connectivity Analyser*" tool is demonstrated with an illustrative example in section 5.4

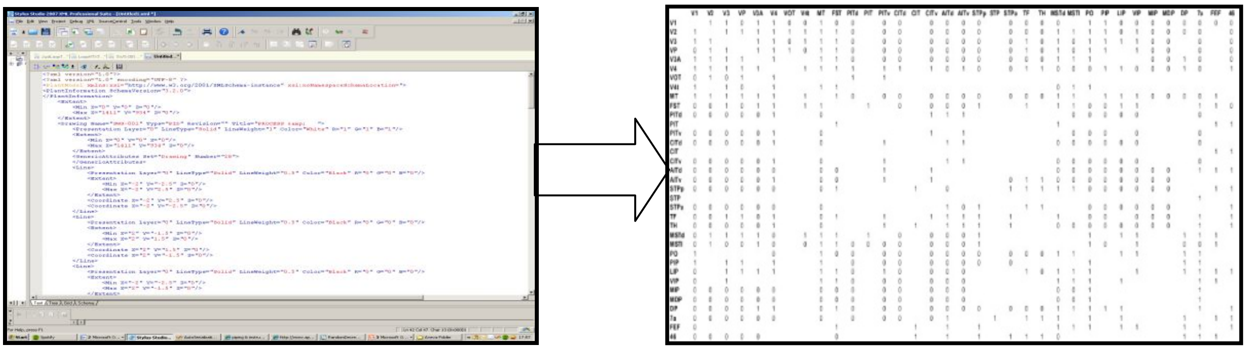


Figure 48: An example of connectivity matrix (on the right hand side) from process XML description

3.10 CAEX Compliant XML Output

CAEX is another relevant schema for encoding and structuring vital engineering data locked-up in documents generated during the design and construction of process plants. CAEX was considered as an alternative to ISO15926 at the early stage of the PhD work, hence its inclusion in the thesis. CAEX is an XML schema for mark up declaration. It is a platform and vendor independent object-oriented data model for machine information exchange and storage.

3.10.1 Background

CAEX started as a university project at the RWTH Aachen at the chair of process control engineering (Epple, *et al*., 2002) with the industrial support of the ABB corporate research Ladenburg and developed by a committee of ABB engineers and others (Fedai and Drath, 2005) In 2004, CAEX was published as part of the DIN V 44366. After a positive international voting, CAEX has been published as part of the IEC PAS 62424 in May 2005. In 2007, the next IEC standardization step has successfully been passed, it is published as IEC 62424 CDV -committee draft for voting (Schleipen *et al.,*2008)

CAEX defines structures for the definition and storage of objects with their characteristics and relationships. CAEX supports object oriented concepts such as encapsulation, classes, class libraries, instances, instance hierarchies, inheritance, relations, attributes and interfaces in storing vendor independent hierarchical object information. CAEX supports three types of classes and corresponding libraries:

System Unit Classes: describe physical or logical plant objects or units including their technical realization and internal architecture. *SystemUnitClasses* are collected in libraries of the type *SystemUniClassLib.*

Role Classes: these are abstractions of concrete technical realization of physical or logical plant objects. *RoleClasses* do not describe the concrete internal implementation of the object. It is used in order to define the requirements for a plant catalogues. They are packaged in *RoleClassLib.*

Interface Classes: describe types of interfaces. *InterfaceClasses* comprise a set of specific attributes used for specifying interfaces for example, *RoleClasses* and *SystemUnitClasses*. They are required in order to define relations between objects. *InterfaceClasses* are packaged in *InterfaceClassLib.*

XML that conforms to the CAEX schema can be constructed with the aid of an XML visualisation tool such as Stylus Studio. The XML documents can also be created from engineering drawings by computer-aided engineering (CAE) tools such as ComosPT® from Innotec®.

3.10.2 CAEX Plant Items Specification

Figure 49 specifies all the necessary elements needed to completely describe items and connections, in a CAEX format, in a typical process P&ID. The following terms immediately become evident and useful in the process topology specifications:

Product Connections: symbolize the coupling of two pieces of equipment with the possibility of material transfer between them (pipe-pipe, pipe-vessel).

Process Connection Line: symbolizes the information flow from the control world to the physical process or vice versa. The *ProcessConnectionLine* symbolizes the functional coupling between a PCE request and the material balance point, but not the actual layout in the plant

Signal Line: symbolizes the functional influence between PCE requests, and not electrical wiring.

Interfaces (external and internal): external interfaces are CAEX means to describe product flow (or signal flow) between units while internal interfaces describe flows within a unit.

Standards: Pertinent standards relating to CAEX include: ISO 10628, which specifies the general rules for flow diagrams for process plants and IEC PAS 62424, a specification for PCE requests in P&ID, data exchange between P&ID tools and PCE-CAE tools. These standards are discussed in Section 3.4.3.

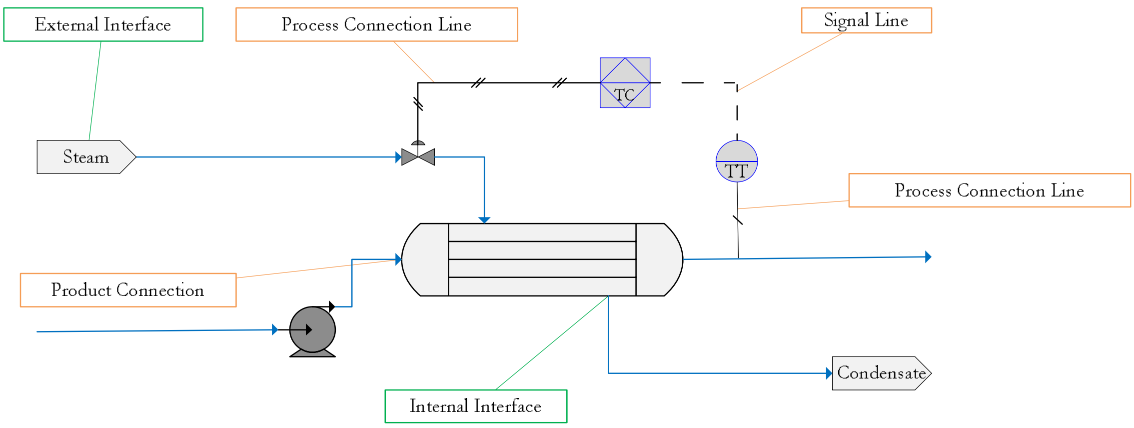


Figure 49: CAEX specification of process P&ID.

3.10.3 XML Encoding of Plant Items

Figure 50: External references in a CAEX complaint XML file shows the header of a CAEX complainant XML file as depicted in Figure 50. The header contains references to external libraries which are in turn compliant with the referenced international standards: ISO 10628 and IEC 62424. This indicates that new libraries need not be defined for each and everything in CAEX. CAEX acts as a backbone that integrates existing libraries. CAEX explicitly supports accessing external files by means of the CAEX element *ExternalReference*. The definition of *ExternalReference* comprises relative path such as *".\ISO10628-2007.xml"* as shown in Figure 50 and an alias name "*ISO10628*" that allows for internal access to the external file. In addition to the standards, external CAEX files can be referenced as well.

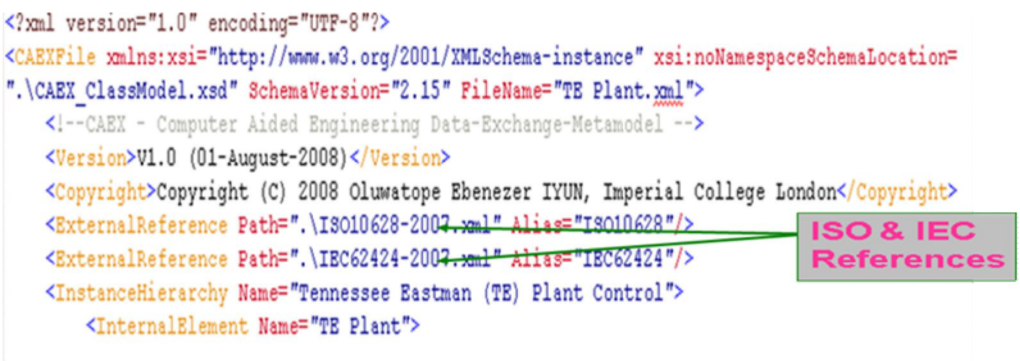


Figure 50: External references in a CAEX complaint XML file

The ability to reference remote CAEX files allows for smooth work flow and project integration. Examples of external interface description are shown in Figure 51 as *Off Sheets A*, *D, E, C, P* and *W* which represent process streams external to the current P&ID. Examples include a feed stream from a feed pre-processing unit or storage tank, product stream flow or effluent/waste stream flow.

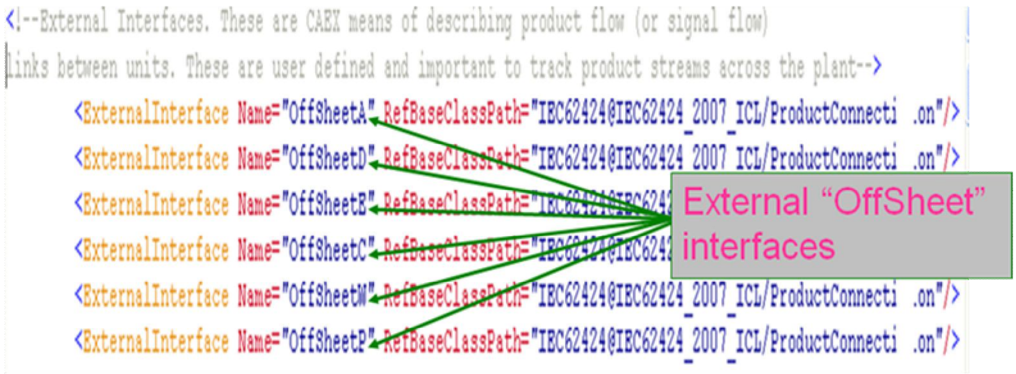


Figure 51: CAEX external interfaces

List plant elements extracted from a process P&ID is shown in Figure 52. The list of plant items includes equipment such as a pump, controllers, indicators, pipes and piping components. For example, "TC007" in Figure 52 designates the temperature controller in loop 007

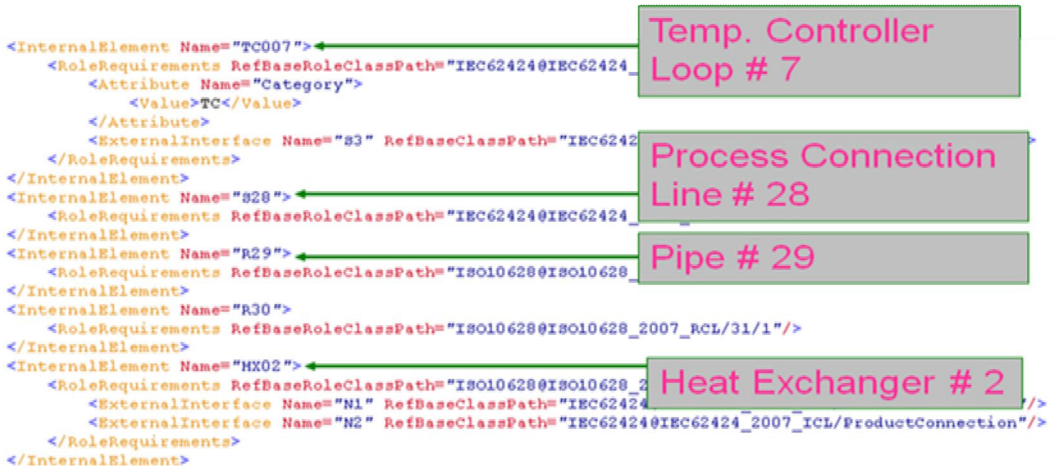


Figure 52: Output from Stylus Studio® - Components

The symbol "S" is used to represent process connection line while "R" is used to represent a pipe. "N" is used to represent product connection interface. Some of the naming conventions are listed in Table 10.

Table 10: An example of naming convention for encoding process plant items

Symbol	Designation
Interfaces	
S_x	Signal Line Interface
P_x	Process Connection Interface
N_x	Product Connection Interface e.g. Nozzle
Connection lines	
S_{xx}	Process Connection Line
S_{xxx}	Signal Line
x = integer number	
Components	
V	Valve
R	Pipe
CM	Compressor
A	Analyzer
HS	Heat Source
HZ	Heat Sink
T	Pipe joint
Unit equipment	
HX	Heat Exchanger
RX	Reactor
CD	Condenser
SP	Separator

Plant topology refers to unit's entities such as equipment configuration, the connectivity among them and external elements or surrounding environment. The process topology descriptions are usually available in such forms as P&ID and other engineering drawings. Process schematics become more useful when they are available in electronic format so that computational manipulations such as search to be performed on the items process schematic to show the sequence of event such disturbance propagation away from a source.

Figure 53 shows a schematic of a processing equipment unit *RX01* and its corresponding XML description to its right. The figure shows that process connection line *S22* is connected to the reactor, *RX01* and level controller, *LC003*. The interface between *S22* and *RX01* is *P1*. *S22* is connected to *LC003* through *P2* interface. Signal line *S922* is connected to *LC003* through a common interface *S1*.The complete XML description of the unit and the connectivity is shown in the right hand panel of Figure 53.

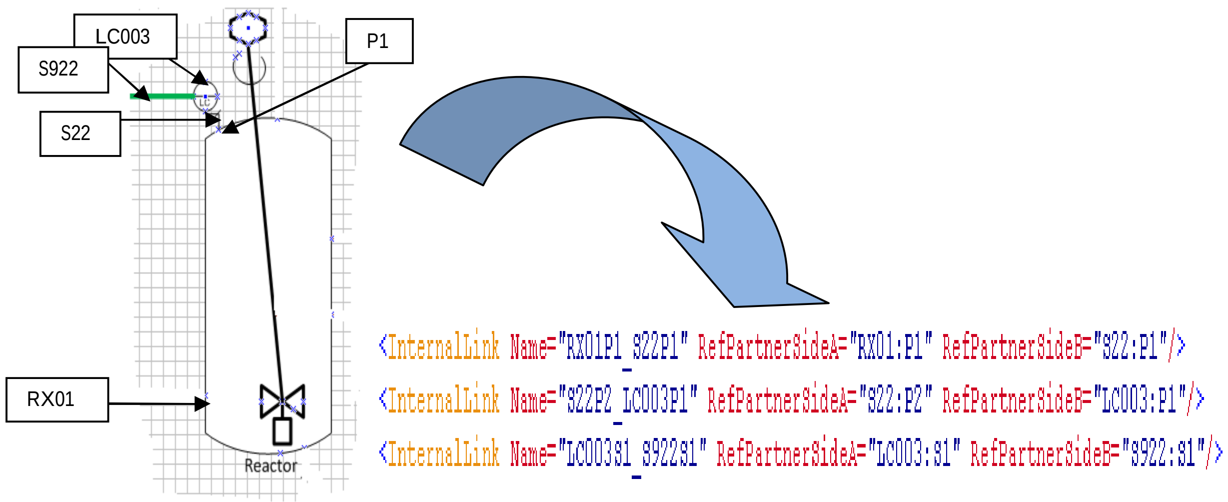


Figure 53: A process unit with its corresponding CAEX XML description.

The interfaces defined and used in the XML description can be filtered out during the parsing of the XML file in the final connectivity analysis, leaving only the components and connectivity information. For example, the XML description:

"<InternalLink Name="RX01P1_S22P1" RefPartnerSideA="RX01:P1" RefPartnerSideB="S22:P1"/> "

will produce "RX01--->S22" as directional link between *RX01* and *S22* without showing the interface *P1*.

3.11 Chapter Summary

This chapter has presented various techniques for accessing process topology data. Relevant standards have been reviewed in the context of the thesis requirements. The chapter discussed commercial software tools available and a selection was made based on the project requirements. The chapter also introduced XML exports of process plant topology.

4 Design and Implementation of Process Connectivity Analyser Tool

A substantial part of the work reported in this thesis involves engineering research findings into a software tool with familiar and friendly user interface. This is a key project deliverable as required by the project's sponsor. A substantial amount of time and effort was invested in the software engineering process. This chapter captures the essential details the software development process. Activities carried out include requirement analysis, software design, coding, testing and deployment.

4.1 Software Development Methodology

Requirements specifications for most research projects are subject to continuous changes and modification as new discoveries unfold. Consequently, the appropriate software practice would be the Agile software development methodology (AgileManifesto.org, 2011). This approach was adopted in the current work as it allows for continuous software development and modification to existing requirements based on interaction with project's industrial sponsor throughout the course of the research project.

For implementation, an *objected oriented* software development paradigm was adopted. The *object oriented* approach to software development allows complex programming problems to be broken down to small manageable modules (objects) that are easy to understand, maintain and keep bug-free(Medvidovic *et al*., 2002 and Warnars, 2011). Co-operation among the various resulting objects allows complex problems to be solved. For example, a crude oil refinery can be broken down into various constituent objects as distillation columns, pipes, equipments and instruments. Each of these components can be broken down further if necessary, developed, and tested independently before integration.

Object oriented approach to software development supports the following:

- *Classes:* these are the abstract template for creating similar objects. For example, a generic equipment class can be used to create specific equipment objects such as a pump with a unique tag name.
- *Encapsulation:* allows internal operations of a class functionality to be hidden away from the programmer. Interaction with the class is achieved through a set of well defined, publicly available interfaces.
- *Abstraction*: this is the process of capturing and representing only relevant simplified data structure of real-world objects. For example, a man can be represented by just his name, age and date of birth leaving out other details such as height, colour of his hair, profession and so on.
- *Inheritance*: allows one class (the *sub-class*) to be based upon (inherited from) another (the *super-class*) and inherit all of its functionality automatically. Modifications to the *sub-class* are permitted to create a more specialised version of the class.
- *Polymorphism:* this process describes the ability for an object to change its behaviour according to how it is being used in the program. This allows identical naming conventions which are distinguishable by the parameters that are passed into the functionality.
- *Modularity:* object-oriented programming also permits increased *modularity* where individual classes or groups of integrated classes can be thought of as a module of code. *Modularity* allows written codes to be re-used in other software projects thus reducing developmental efforts, time and cost.

4.2 Requirements Analysis and Specification

This section describes the pertinent requirements analysis and specification for the research project. The project went through continuous iterative changes in terms of functional requirements and implementation. For example, the project started with IEC 62424 standards using CAEX for encoding process plant schematics and connectivity information and later changed to ISO 15926 standards.

The bulk of the requirements have been tabulated Table 11 to show the position of each feature/requirement in priority stack. Each feature has been assigned a degree of importance as mandatory, desirable or optional.

The end product of the research efforts is a software tool in Windows environment (Microsoft .Net) that enhances process control engineers' ability to perform plant-wide diagnosis and performance problems using process data, directionality and connectivity information from process topology such as P&ID and process understanding (know-how).

Table 11: Requirements analysis

ID	Feature	Functional Specification	Implementation (In Visual C #)	Priority	Notes / Comments
R01	Load and parse ISO 15926 compliant XML file	Create an *n* x *n* connectivity matrix where n is the number of parsed items	File read from storage medium, pre-processed and tagged-items extracted	Mandatory	Implemented
R02	Handle multiple XML files from a number of P&ID exports	Combine multiple well-formed XML files into a single well-formed XML file	File multi-select function implemented, each file is read, filtered, defragmented and recombined as an entity	Mandatory	Implemented
R03	Provides insights about cause-and – effect analysis from data- driven analysis	Linkage of connectivity information with results from process signal analysis	-defines data format - load and parse data	Mandatory	Implemented
R04	Incorporate knowledge about the process	Hard coded logic in C # or Use of Expert System Shell / Programming	Logic coding of Physics and expert knowledge via interviews and discussions on site	Mandatory	Implemented
R05	Enable export of Connectivity Matrix	Integrate Excel application within the tool	Import and Instantiate Excel application. Copy entries in connectivity matrix to instantiated Excel application	Desirable	Implemented
R06	Path tracing	User specifies a starting point and an end point	Breadth-First search of the connectivity matrix	Desirable	Implemented

		to check for feasible path(s)			
R07	Display results showing process variables (controllers and measurement points) with option to include process equipment and other plant items	Implement a filter to remove or include process equipment	Provide users with a check box to facilitate selection of option	Desirable	Implemented
R08	Remote access	Convert to Browser application	Modify WPF XAML to XBAP	Optional	Implementation requires slight modification
R09	Distinguish between process lines and utilities lines	Interpret information embedded in tagged item	Specify correct labelling during P&ID drawing	Optional	Not required in the current project. Implementation with modification
R10	Integrates with existing tools	Define interfaces that seamlessly integrate with other software tools	Work closely with vendors	Optional	Requires collaboration with software vendor(s)

4.3 Software Tool Design

In order to manage and organise large programming efforts in the research, classes were design to encapsulate essential elements of the software functionality. Co-operation and interaction among various classes and the objects created by these classes allow required tasks to be accomplished.

4.3.1 Main Classes

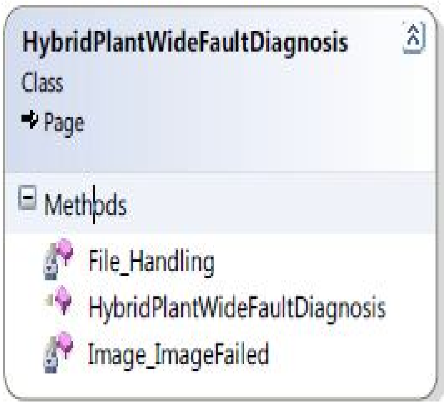
This section describes main classes designed and implemented in *Process Connectivity Analyser* The classes are fundamental to software implementation in object-oriented environment and were developed specifically for the PhD project.

Each class representation shows the class methods, fields as well as properties where implemented.

One of the most important classes is the *IDtoTags* class whose component field, properties and methods are shown below after *HybridPlantWideFaultDiagnosis* class. The *IDtoTags* class is essentially an XML parser that loop through the XML file to extract useful elements. The parsing algorithm is depicted in Figure 54.

The algorithm checks for the existence of an XML file before extracting relevant tags. The algorithm extracts, sorts and stores tagged items under one category as a control loop, equipment or instrument in a re-sizable data structure. Every connection to a process equipment is via a nozzle either explicitly drawn on the P&ID or automatically inserted as a *virtual nozzle* when not specified/drawn on the P&ID. In either case, the nozzle is not included in the final parsed items as they are not required for the PhD project.

HybridPlantWideFaultDiagnosis Class



This class is responsible for all file handling. It reads files from storage location on the local machine or any portable drive. *HybridPlantWideFaultDiagnosis* class also implements all error checking logic associated with file reading and processing. For example, any attempt to read a non-existing file or file with incorrect format.

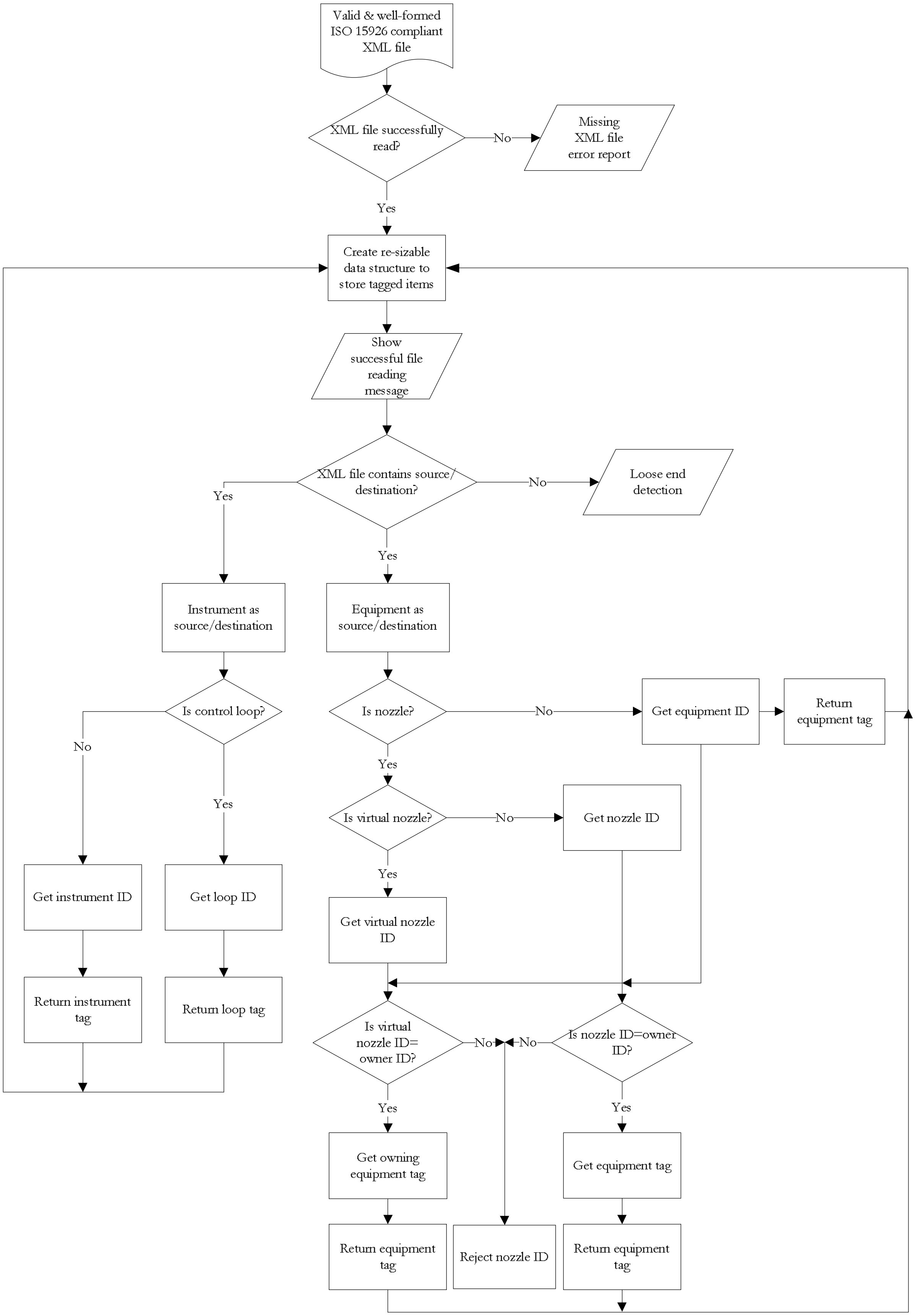
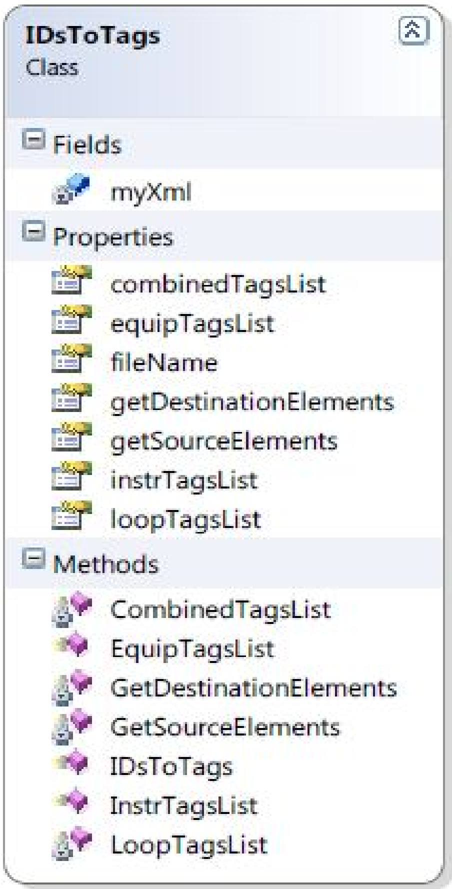


Figure 54: Flowchart for ISO 15926 XML parsing algorithm

IDsToTags Class

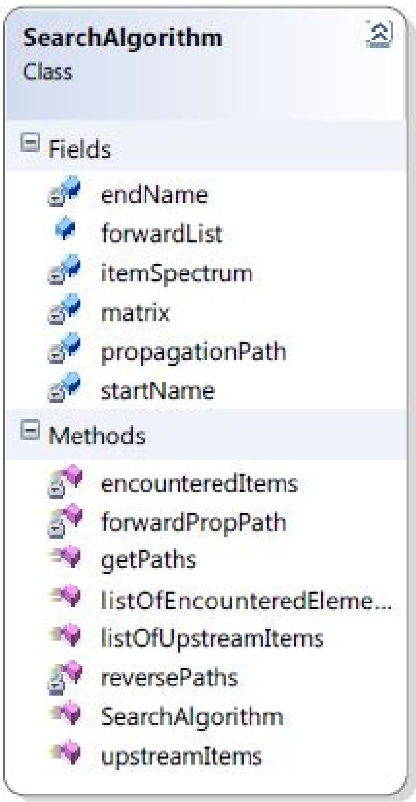


IDsToTags is the workhorse of the *Process Connectivity Analyser* tool. It parses the XML file once correctly loaded to extract plant items and their directional connections.

The class defines some methods (functions), a field and a number of properties to expose the field which is kept private to the class and accessed via get and set properties. Properties provide a level of abstraction which allows the field to be changed without affecting the external way they are accessed by other classes.

However, properties are not always required to encapsulate the field. The properties can return values on their own accord.

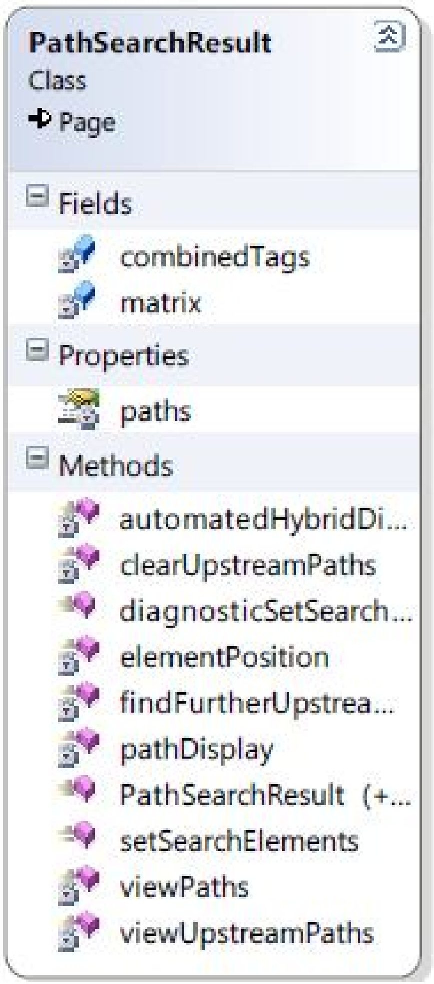
SearchAlgorithm Class



SearcchAlgorithm class manipulates the connectivity matrix and implement a depth-first search algorithm to find physical path from a chosen starting element to the end element.

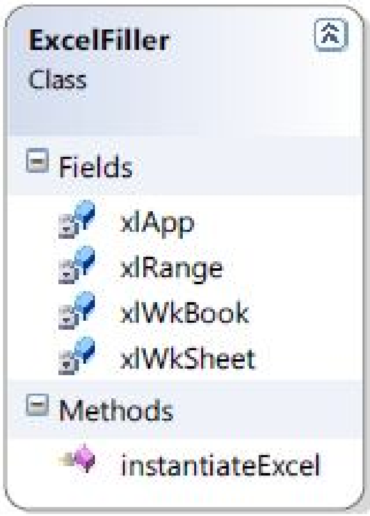
The classes described above interact and co-operate to perform the various tasks set out in Table 11. Figure 55 shows how the classes written in this project reference external libraries. The external libraries are resources packaged within Microsoft Visual Studio's integrated development environment as part of .NET framework 3.5 technology to perform routine low level functions such as reading and writing of data to streams so that developers can concentrate the actual task of software development

PathSearchResult Class



This class is responsible for pre-processing and display of search results. Pre-processing includes such operations as filtering, merging and removal of redundant elements. Display involves data-binding of items to view to windows presentation foundation (WPF) controls.

ExcelFiller Class



This class handles calls to Microsoft Excel® spreadsheet application and filling the spreadsheets with the connectivity matrix already created by the *ConnectivityMatrixGenerator* class

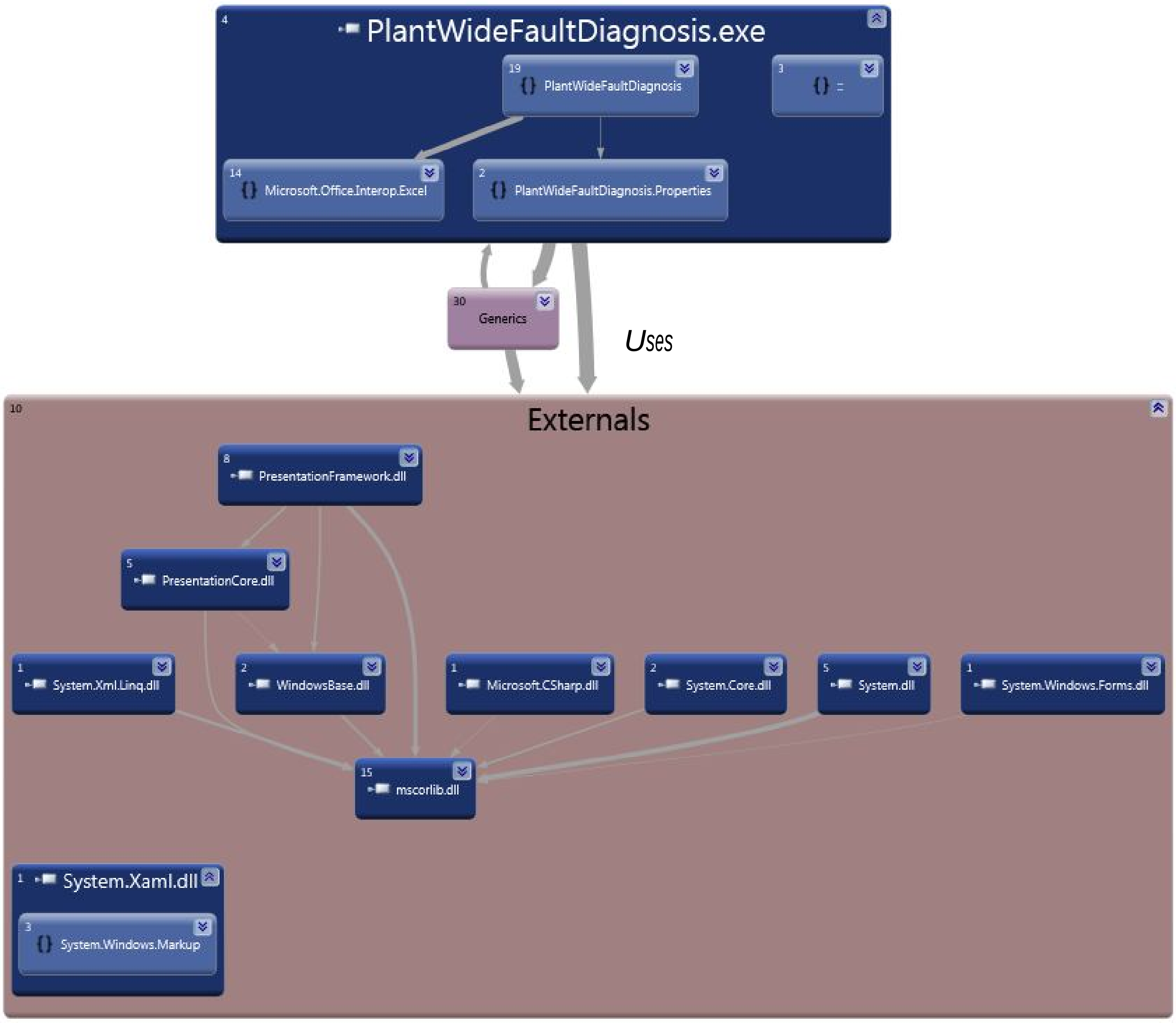


Figure 55: Reference to .NET framework class libraries

Detailed interaction among classes written in the project is depicted schematically in Figure 56.

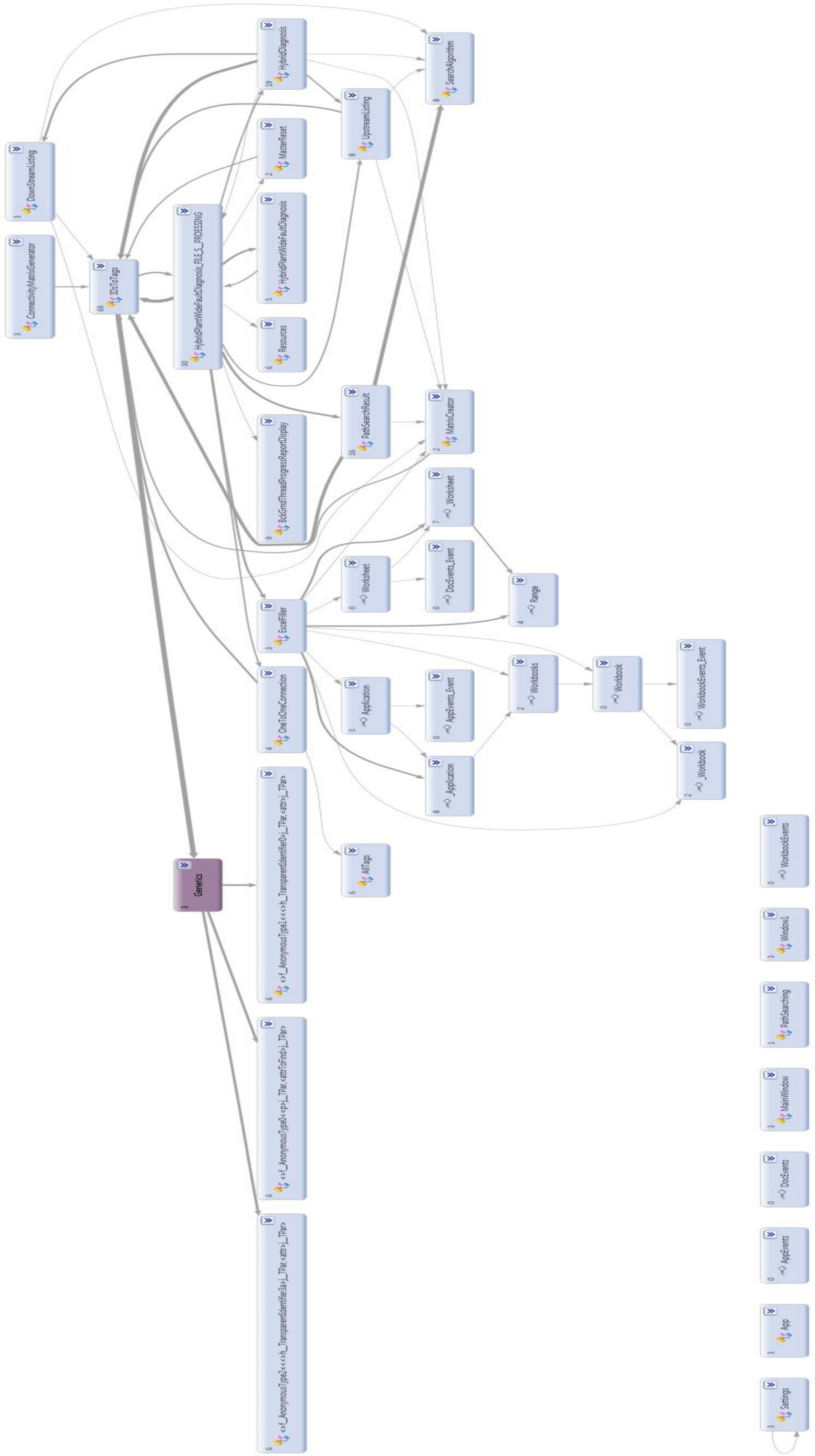


Figure 56: Classes interaction automatically generated with Visual Studio 2010 software development tool

Figure 56: Classes interaction automatically generated with Visual Studio 2010 software development tool

4.4 Requirements for a Standalone Software for Creating and exporting Process Connectivity Matrix- A Stripped-down Version of the Final Software tool

This section describes the requirements for a standalone software component, a lightweight version of the overall software tool that automatically converts XML exports of electronic process piping and instrumentation diagram (P&ID) into a connectivity matrix. The connectivity matrix produced can be exported to Microsoft Excel application as a standalone application or processed further. Automated creation of a connectivity matrix is given special consideration here due to its central importance to the project. The software described here functions as a standalone tool capable of exporting process connectivity matrix to Microsoft Excel application.

The reason for the inclusion of this lightweight version of the software under a separate heading is that the software was used by two MSc students who carried out rigorous and systematic testing on the software against a set of requirements. The results of these tests are presented in Section 4.6.

4.4.1 Purpose and Scope

The main purpose of this section is to identify the requirements and constraints that the component software tool required for automatic conversion of process description in XML to connectivity matrix must satisfy. The proposed component software should be designed, implemented and tested under various scenarios with sample process P&ID to ensure that it meets all the requirements. It should provide interface that allows user to upload an electronic copy of a P&ID XML file. The parsed file should generate a connectivity matrix for further processing by the tool software in Microsoft .NET environment. The requirements of the component software tool will be considered under the following categories: functional requirements, interface requirements, software integration requirements, converted files integration, directionality requirements and quality attribute requirements.

4.4.2 User Characteristics

The user of the final software tool will be a process plant operator who will not necessarily be a computer expert. The user should not be bogged down by trying to understand how the tool works but should be easier for the operator to use and understand and analyse the result produced.

4.4.3 Dependencies

The project that will make use of the component software and connectivity matrix output will be developed in Microsoft .NET environment. Integration and compatibility of the technology used in this work with Microsoft .NET technologies will be an essential requirement.

4.4.4 Constraints

AutoCAD DWG files contain drawings of structures like tanks. These lines can look like signal lines or pipelines. Attempt should be made to use tag naming to identify the entity type on the P&ID. This is usually based on the ISA 5.1 standard. However, not all P&IDs comply with the standard. Typically, the ISA standard is used with some minor variations for the instrumentation symbols in almost every P&ID. There is usually a first page on the set of P&ID symbols which normally gives the legend for the drawing and is mostly similar to the ISA standard.

4.4.5 Requirements

This section describes the requirements and attributes of the AutoCAD conversion software for generating a connectivity matrix from XML description of the process schematics.

Functional Requirements (FRQ)

FRQ 1:

ID	FRQ1
PURPOSE: Uploading XML file	
PRIORITY: HIGH	
DESCRIPTION: The software tool should allow the user to browse to and upload electronic process P&ID in XML format structured according to ISO 15926 standard. Facility must be provided for loading a single or multiple XML files in batches for the purpose of parsing the file(s) and generating connectivity matrix of specified elements in the drawing. The user should be able to do this through a user-friendly graphical user interface (GUI).	
INPUT: XML file compliant with ISO15926 standard	
OUTPUT: The user should receive a confirmation that the file has been successfully loaded, otherwise and error message should appear, indicating what went wrong.	

FRQ 2:

ID	FRQ 2
PURPOSE: Deletion and modification of uploaded file	
PRIORITY: MEDIUM	
DESCRIPTION: User should have a means of removing / deleting an uploaded file with a warning message to the user before deletion is effected. The user should be able to re-arrange the order in which uploaded files are stored and eventually converted.	
INPUT: XML file compliant with ISO15926 standard	
OUTPUT: The user should receive a confirmation that the file has been successfully loaded, otherwise an error message should appear, indicating what went wrong.	

FRQ 3:

ID	FRQ 3
PURPOSE: XML files merging	
PRIORITY: HIGH	
DESCRIPTION: XML files exported from process P&IDs of a plant could contain several batches of files and spread over several sheets of paper when printed out. These files/sheets are linked by off-sheet labels. The software tool should be able to link the files together and produce a unified connectivity matrix for different file batches using the off-sheet labels as the interface. E.g. a pipe should run continuously from one sheet to the other until it reaches a target.	
INPUT: Several batches of XML files from P&ID drawings with interface for connection.	
OUTPUT: Integrated file and ultimately, a unified connectivity matrix representing the entire plant entity.	

FRQ 4:

ID	FRQ 4
PURPOSE: Directionality	
PRIORITY: HIGH	
DESCRIPTION: Flows in pipes and signals from control instruments follow a specific direction. This information should be inherent in the connectivity matrix formed. For example the order in which elements are connected should be preserved. It should be obvious if Unit A is directly connected to Unit B and not the reverse. This should be reflected in the connectivity matrix formed.	
INPUT: Integrated XML file from process schematic that conforms to ISO 15926 standards.	
OUTPUT: Unified connectivity matrix representing the entire plant entity with directionality information captured.	

Interface Requirements (IRQ)

ID	IRQ
PURPOSE: Usability	
PRIORITY: NORMAL	
DESCRIPTION: Here, simplicity is the key word. Proposed tool should be user-friendly and provide help functions. The user should not spend a lot of time figuring out how to use the tool e.g. looking for the 'browse' button to upload a file. It should be easy to learn and use.	
SUGGESTED FUNCTIONALITY: Help facility should be provided that will help a novice or confused user.	
RELEVANCE: The tool is targeted at both computer experts and non-experts alike and as such should be usable by any user independent of the user's computing knowledge.	

Software Integration (SIRQ)

ID	SIRQ
PURPOSE: Integration with other software	
PRIORITY: NORMAL	
DESCRIPTION: The tool will be an integral part of a larger ongoing project and should therefore be easily integrated into the final tool. The final tool will be developed in a Microsoft .NET IDE so it is essential that this tool comply with the requirements for such integration to take place.	
ACHIEVABLE: Microsoft .NET platform integration of multiple programs written in various high level languages.	
RELEVANCE: A unified software tool is the overall goal of the entire project. Therefore individual component software developed along the line must be incorporated into the larger project.	

Reliability and Performance Requirement (RPRQ)

ID	**RPRQ**
PURPOSE: Robustness and high throughput **PRIORITY:** HIGH **DESCRIPTION:** The tool should be very stable and process input data within a reasonable period of time. **ACHIEVABLE:** In addition to other requirements, reliability and performance characteristics of the tool will be directly tied to the overall quality of the software tool. Backup and recovery facilities should be provided to avoid any loss of valuable data in case of system failure. **RELEVANCE:** In order to ensure a high degree of confidence in the tool, a highly efficient and stable software is much desirable than a crashing one. Also users expect output within a reasonable period of time than wait for eternity.	

Maintainability and Evolution Requirements

The tool should be easy to maintain by providing adequate documentation and well commented code. The software should be able to adapt and incorporate additional functionality as the need arises in the future.

4.5 Implementation

This section describes the technology (.NET framework), software development tools (Microsoft Visual Studio 2010) employed and the high level programming language (C #) used in coding the software requirements.

4.5.1 .NET Framework

The .NET framework technology (Microsoft, 2011) is the Microsoft® initiative that allows programmers to develop applications in different languages that can easily interoperate. As shown in Figure 57, application can be developed using common high level programming languages such as C++, C#, Visual Basic, Java and C and pre-processed into intermediate languages and finally to Common Language Runtime that native processor can manipulate. This allows for flexibility and interoperability in application development. The common language runtime ensures portability on various computer systems. As at the time of writing this thesis, the latest version of .NET framework according to Microsoft website is .NET framework 4.

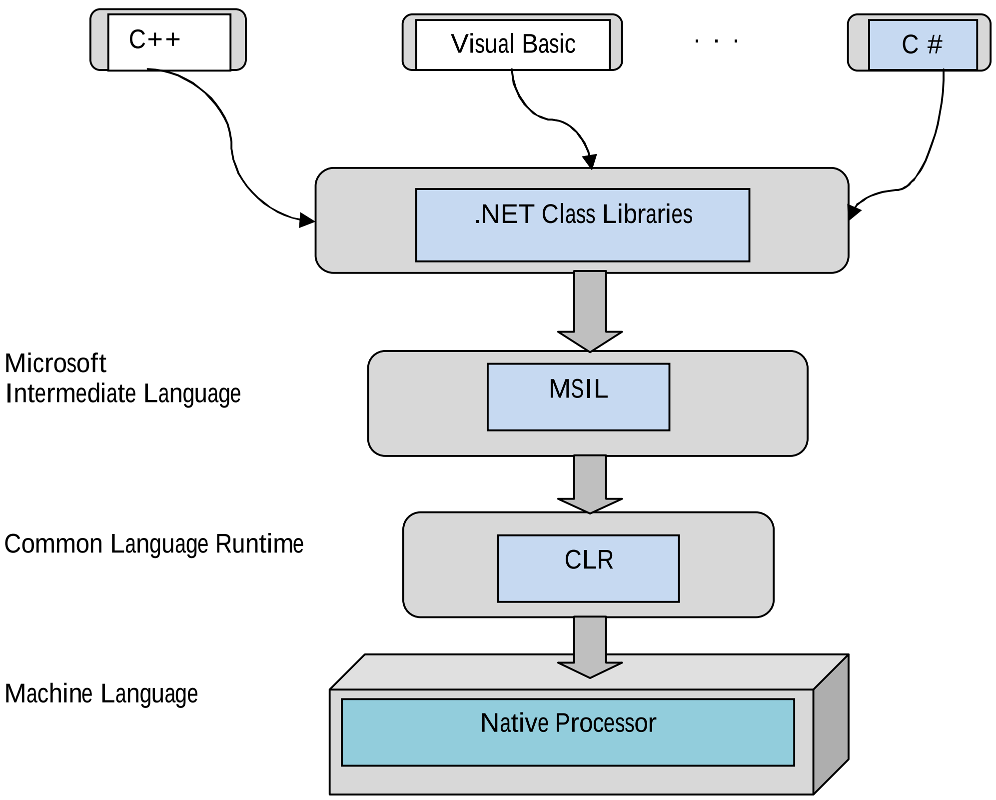


Figure 57: Microsoft .NET technology framework layers

The core components of .NET framework technology are:

- *.NET Class Libraries*: a library of classes, interfaces, and value types that provides access to system functionality and is designed to be the foundation on which .NET Framework applications, components, and controls are built.

- *Microsoft Intermediate Language:* This is a language-independent and Central Processing Unit (CPU)-independent representation of the code written in high level language such as C # after compilation

- *Common Language Runtime*: The .NET Framework provides a run-time environment which runs the code and provides services that make the development process easier.

4.5.2 Visual Studio

The Microsoft Visual Studio is an integrated development environment (IDE) that provides code editor, designer and compiler for developer’s chosen high level development language such as C #, Visual Basic or C++.

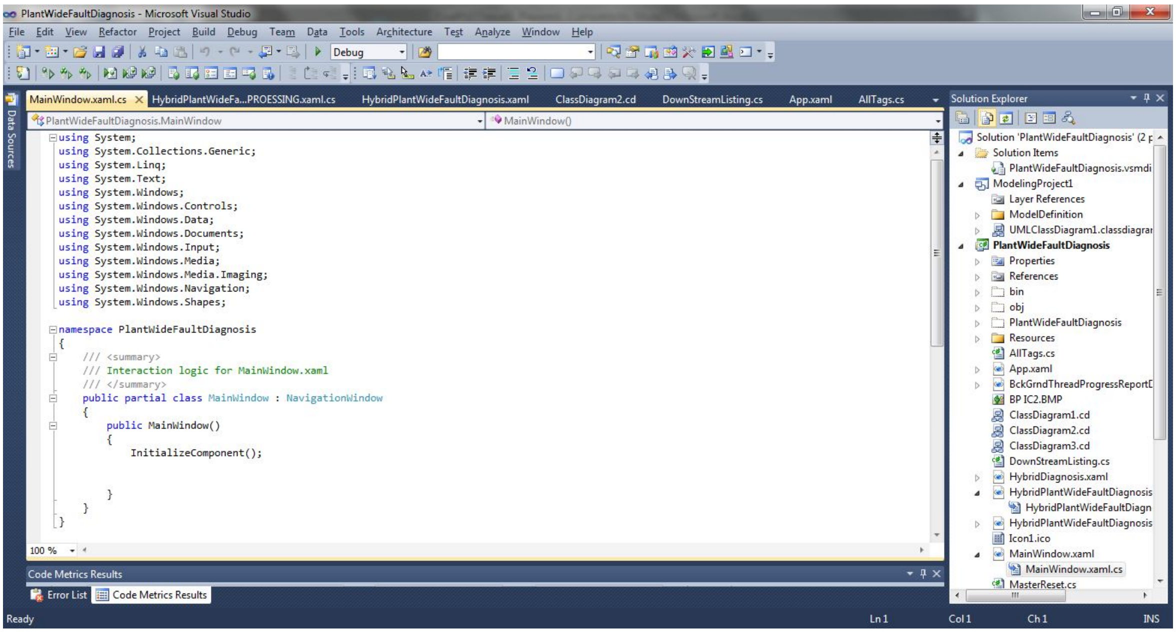


Figure 58: Microsoft Visual Studio 2010 integrated development environment (IDE) user interface

4.5.3 The C # Programming Language

Software developed from the research work utilized the Microsoft® .NET technology using C# as the core programming language to integrate various development components. C# is a powerful, object based and fully integrated into the .NET technology. Applications developed using C# meet the core requirements of modern applications. That is they are scalable, distributed, web-based and secure.

While the bulk of the code was written in C #, other .NET technologies were also utilized in the software development. These set of technologies include:

- Language integrated query (LINQ): LINQ allows for easy querying and updating of data, in memory and from external sources. This essentially allows for an innovative way of querying against data as well as objects. It uses the standard select, from, where syntax for querying data or objects.

- Windows presentation foundation (WPF) and extensible application mark-up language (XAML): XAML is a declarative mark-up language. As applied to the .NET Framework programming model, XAML simplifies creating a user interface for a .NET Framework application. This can be written in text using an XML description. An example of XAML written to create a button user interface element is:

```
<Button>
        <Button.Background>
                <SolidColorBrush Color="Blue"/>
        </Button.Background>

        <Button.Foreground>
                <SolidColorBrush Color="Red"/>
        </Button.Foreground>

        <Button.Content>
                        OK Button!
        </Button.Content>
</Button>
```

4.6 Software Testing

The software was subjected to comprehensive and systematic testing by two MSc students who actually used the lighter version of the software in their research work. The feedback is generally good as the software meets all the functional requirements. There are however some suggestions for improvement on a number of non-functional requirements such as its appearance. Table 12 and Table 13 present the summary of the structured tests carried out by the testers. It should be noted that each tester worked on a different project and despite the variability, the universality of the software have been demonstrated as each tester was able to find the software fit for the purpose of his respective project needs.

Table 12: Summary of software testing by tester #1

Software:	**Connectivity Matrix Export to Excel Application**			
Test Date:	20th August & 28th September 2010			
Tester:	David Babarinde Alabi			
TEST 1				
Screen	**Actions taken**	**Settings**	**Results and bugs**	**Comments**
User Machine	Start the Software (Executable)	Double click on the executable icon	OK, worked normally	
1	Press "START XML PROCESSING"		OK, worked normally	
2	File Pre-processing-merging, filtering etc if necessary for Multiple P&IDs	Click on the appropriate Menu Item and sub-menu to choose files to process	N/A	
2	Multiple files selection	Press down "CTRL" button to select multiple XML files for pre-	OK, worked normally	

		processing		
2	Save final processed files on local machine hard drive (C:/)	Choose this option as submenu	OK, worked normally	
2	Press OK to confirm operation		OK	
2	Save final processed files on removable storage device such as USB flash pen to get around "Write Permission" restriction	Choose this option as submenu	OK	
	Pressed OK to confirm operation		OK	
2	Exit the operation without selecting file or files	Software response	No bug, exits normally	
2	Loading XML file to work with	Click on the appropriate Menu Item and sub-menu to choose appropriate XML File	OK	
2	Multiple files selection	Press down "CTRL" button to select multiple XML files for pre-processing	OK	
2	Press OK to confirm operation	Program response	OK	
2	Pressed OK to confirm operation		OK	
2	Exit the operation without selecting file or files	Program response	Exits normally, no bug	
TEST 2				
Screen	Actions taken	Settings	Results and bugs	Comments
2	Observe Output on screen 2 after file selection	All tag items are displayed	OK	

2	Click "Export ConnectivityMatrix to Excel" button	The response depends on the size of the XML file	OK	
2	Click "View 1:1 Connections" button		OK	It could be nicer if the scroll bar for LHS display can be placed at the middle in order to move the two parts together
3	Scroll down the list displayed on screen 2		OK	
3	Click the back button		OK	
2	Click the back button		OK	
1	Click the back button		OK	
2	click on the "<<HOME" button		OK	
2	Click on the "<<Help" Menu		OK	The program could be made to display the help file in a new window to display all the texts
1	Close the Window		OK	What about a new button to close the program beside 'Start XML processing' button?
2	Close the Window		OK	
3	Close the Window		OK	
TEST 3				
Screen	**Observation (Non-Functional Testing)**			**Comments**
1,2,3	Software Usability			The software was very suitable for

				what I used it for and it gave me correct results
1,2,3	Software aesthetic value			The aesthetic value is very nice
1,2,3	Response time			Quite OK

Table 13: Summary of software testing by tester #2

Software:	**Connectivity Matrix Export to Excel Application**			
Test Date:	15/09/2010			
Tester:	Giovanni Di Geronimo			
TEST 1				
Screen	**Actions taken**	**Settings**	**Results and bugs**	**Comments**
User Machine	Start the Software (Executable)	Double click on the executable icon	OK, no problem to open the application	Perhaps a logo to recognize easily the program
1	Press "START XML PROCESSING"		OK	
2	File Pre-processing-merging, filtering etc if necessary for Multiple P&IDs	Click on the appropriate Menu Item and sub-menu to choose files to process	OK	
2	Multiple files selection	Press down "CTRL" button to select multiple XML files for pre-processing	OK	

2	Save final processed files on local machine hard drive (C:/)	Choose this option as submenu	Message "DONE"	Could be useful to provide a facility to introduce the destination of the file. Also perhaps in the final message would be useful again show where is the file (like the link in explorer)
2	Press OK to confirm operation			
2	Save final processed files on removable storage device such as USB flash pen to get around "Write Permission" restriction	Choose this option as submenu	OK	
	Pressed OK to confirm operation		OK	Could be useful to provide a facility to introduce the destination of the file. Also perhaps in the final message would be useful again show where is the file (like the link in explorer)
2	Exit the operation without selecting file or files	Software response	Message "You must select files to merge"	
2	Loading XML file to work with	Click on the appropriate Menu Item and sub-menu to choose appropriate XML File	OK	Always open the file in C: it would be good open in the previous folder used. Also is good have the option to open the latest files used
2	Multiple files selection	Press down "CTRL" button to select multiple XML files for pre-processing		
2	Press OK to confirm operation	Program response	OK	

2	Pressed OK to confirm operation		OK	
2	Exit the operation without selecting file or files	Program response	Message "No file has been selected"	The sound is a bit annoying
TEST 2				
Screen	**Actions taken**	**Settings**	**Results and bugs**	**Comments**
2	Observe Output on screen 2 after file selection		OK	I think all the results should be together not in different areas of the window. Useful the identification per category.
2	Click "Export Connectivity Matrix to Excel" button		OK	I have 46 items in my file and it took about ten seconds. I think the report in excel could include extra information such as data about the program, information about the file parsed (P&ID, Project, Area, etc.), summary of the parser (time, number of items). Perhaps this info could be in another sheet of the workbook. Also while the excel file is been created could be good show a sand clock to represent that the program is working. I do not know if the you should fill the excel open to the user. Perhaps you could fill first and then open to the user. If you include the name of the items per column vertical instead of

				horizontal you could reduce the width of the columns and reduce space. Also I found that in some cases is useful do not choose the "0" and leave then blank (perhaps give options for the report to the user). It would be good if the equipments could be organized in the direction of the flow.
2	Click "View 1:1 Connections" button		OK	This window was very useful to check if the connections were OK before export the matrix in excel. I think this button could be an option for the user to check first if the connectivity is corrected before creating the connectivity matrix. However, it is not explicit to understand the columns for sources and destinations, and I do not know if this is important for the user.
3	Scroll down the list displayed on screen 2		OK	
3	Click the back button		OK	
2	Click the back button		OK	
1	Click the back button		OK	
2	click on the "<<HOME" button		OK	
2	Click on the "<<Help" Menu		OK	Perhaps you are already working in the manual of the application. It should be here

1	Close the Window		OK	
2	Close the Window			
3	Close the Window			
TEST 3				
Screen	**Observation (Non-Functional Testing)**			**Comments**
1,2,3	Software Usability			The program was very useful for the project. I could extract the connectivity matrix easily.
1,2,3	Software aesthetic value			In my opinion it looks more like a website. I prefer all in only one windows. I think first the program should show the results in this window and then give the possibility to export in excel. I think that information about the P&ID should be included.
1,2,3	Response time			I had 46 items in my file and it took 10sec to export the matrix to excel.
				Other comments
				I have problems to recognize the control valves. I tried many types and I could not. The only way that I found was adding the valve in the drawing at the end of a pipe. For it was useful to recognize pipes and mixing and splitting points.

4.7 Chapter Summary

Chapter 4 described the design and implementation of the software tool developed in this project. The chapter introduced the technologies and tools employed for the implementation of the software. The chapter presented comprehensive and systematic testing of the software tool results from two testers, who actually used the software in real projects.

5 Software Demonstration

The software tool "*Process Connectivity Analyser"* is the output of research effort reported in this thesis and the aim is to demonstrate the concept of utilizing process connectivity information as contained in P&IDs for operational purposes.

Although the primary aim of the research is to develop software for fault diagnosis purposes using process connectivity information, the tool also offers a variety of functionality besides fault diagnosis. Other operations that can be performed include path tracing from one tag to another, finding plant items upstream of a chosen tag, export of connectivity matrix to Microsoft Excel application among other uses. For example, if a pump was shut off for maintenance, process operators would like to know which portion of the production process would be affected by looking at plant items downstream of such pump. This tool would become handy in such a situation.

This section drills down to the detailed functionality of the "*Process Connectivity Analyser"* tool followed by a worked illustrative example combining data-driven analysis with the connectivity tool for more insight and reasoning.

5.1 System Overview

This software tool takes input data as follows:

- XML file. The XML is generated by exporting an intelligent P&ID, which is traditional P&ID with data repository, of the process to be analysed. The P&ID should conform to ISO-15926 standard.
- Results from statistical data-driven analysis -only needed for hybrid fault diagnosis.

The system is composed of multiple windows with navigation button to move front and back from one window to another in the familiar browser- type of navigation. Results are displayed within the windows depending on user's request and operation.

5.2 System Requirements and Environment

In order to execute the application on a computer system, the following must be in place:

- Microsoft .NET Framework, downloadable free from Microsoft website. As at the time of writing this thesis, the latest version is .NET Framework version 4.
  The framework is available via this Uniform Resource Locator (URL) http://www.microsoft.com/downloads

- Operating system: Windows XP /Windows Vista /Windows 7

- Supported Architectures: x86/x64

- Hardware Requirements: Minimum Pentium 1.5 GHz or higher with 2 GB RAM or more. Minimum 5 GB disk space (for x64)

5.3 Using the Application

This section describes how to start up the application, use controls and perform operations.

5.3.1 Main Operations

Starting up the Application

The application is available as an executable and can be started by double-clicking the application icon. Once the application is running with the main starting window shown in Figure 59, follow the instruction and navigate from one window to another.

Main Start up Window

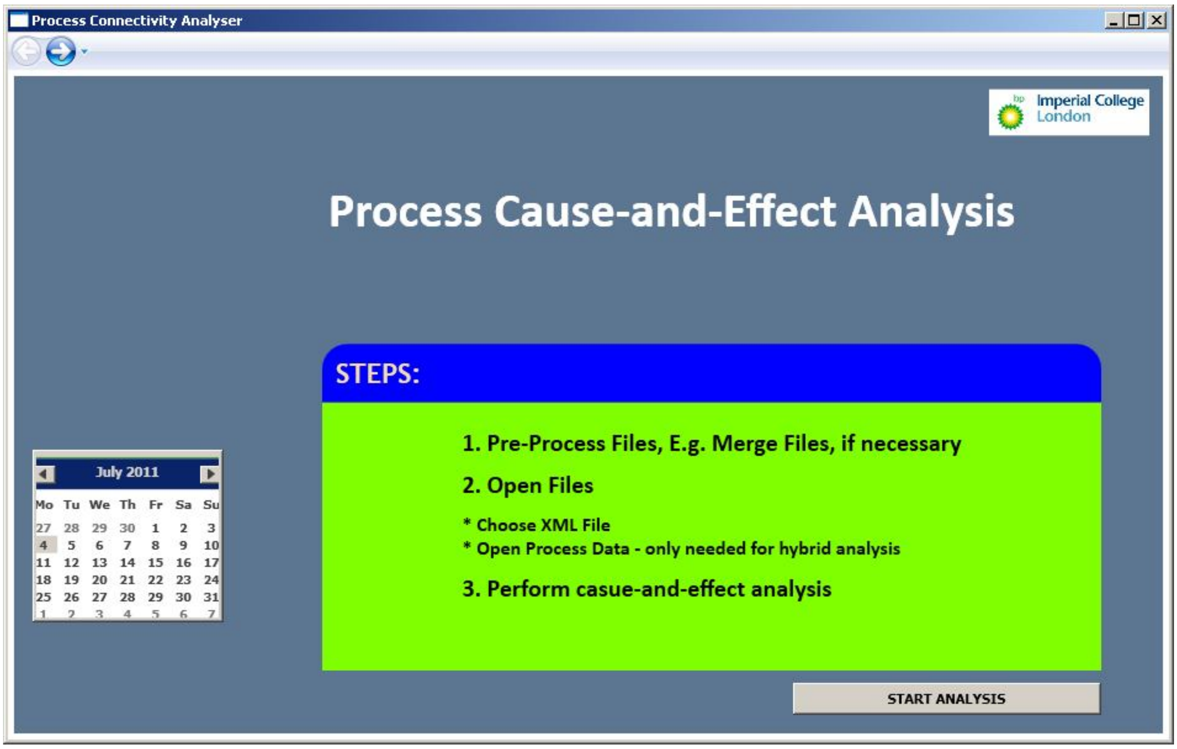


Figure 59: Main start-up window

After reading through some basic required steps, a user will click on START ANALYSIS to continue diagnosis and other analysis. The window also displays a calendar showing the days, month and year. The current date is highlighted on the calendar.

Main Control Panel

Controls for major operations are located on the main control window shown in Figure 60. This allows the user to navigate to other windows depending on the type of operation selected. The various components have been labelled from functionality A to J and each explained subsequently.

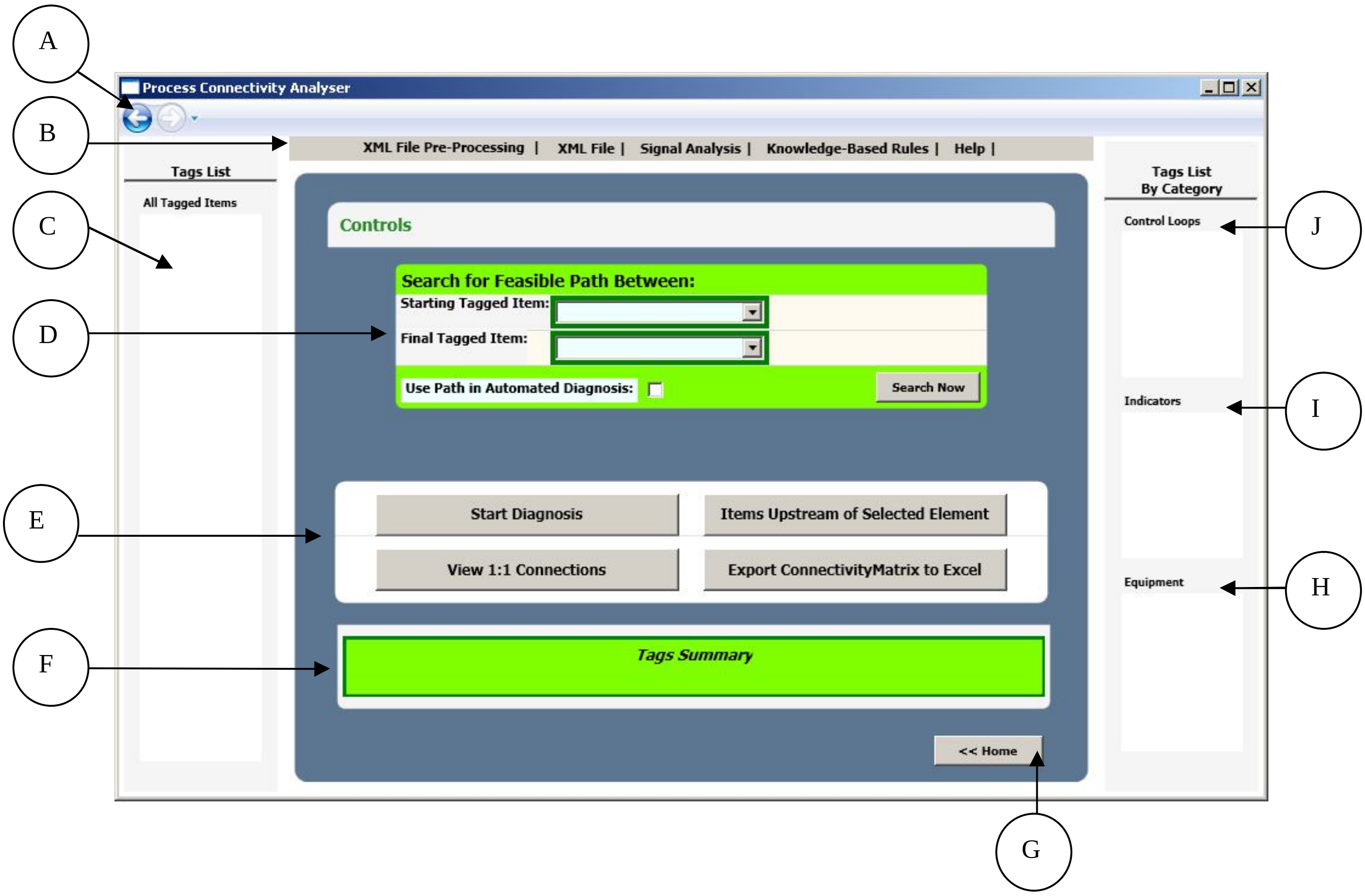


Figure 60: Main control window of the process connectivity tool

(A) Navigation buttons for moving forward and backwards within the application

(B) Menu bar for selecting and uploading file(s). There is also a help menu on the bar as shown below.

- 

For combining multiple XML files from multiple P&IDs. The merged files can be saved on a removable storage device or on the local hard drive of the machine. This functionality is discussed in detail in section 5.3.3.
- 

If no pre-processing, such as files merging, is needed, the XML file can be loaded from any location it is stored.
- 

This allows results from data-driven analysis to be loaded for diagnosis. This can be a plain text such as ASCII characters created using an application such as notepad.

- Knowledge-Based Rules This allows process understanding and knowledge to be checked in reaching a conclusion.
- Help | Online help to aid user's understanding of using the tool more efficiently.

(C) All tagged items list box contains all tags in the P&ID XML export loaded into the tool, sorted in alphabetical order

(D) This sub pane allows the user to choose a starting tag and end tag from the drop down list, to find if a path exists between the chosen elements and how many of such paths exist. To include some diagnostic report, the checkbox must be ticked before clicking *Search Now*

(E) Buttons for major operations, to be described later in the subsequent sections, are located here

(F) This pane produces a summary of tags contained in the P&ID, giving the total lump sum tags and totals based on tag type

(G) Clicking this button takes the user back to the start up screen described earlier

(H) All equipment contained in the P&ID are listed here and sorted in alphabetical order

(I) This box contains indicators from the P&ID's instrumentation, arranged in alphabetical order

(J) Control loops are listed here as contained in the process plant P&ID, sorted in alphabetical order

Search Window

To perform a search operation, a valid XML file must be loaded first. The starting tag and the end tag must be different. The system will prevent the user from performing any illegal operations. Figure 61 shows one of such invalid operation. Here, the user attempted to carry out search operation without loading the process XML file first.

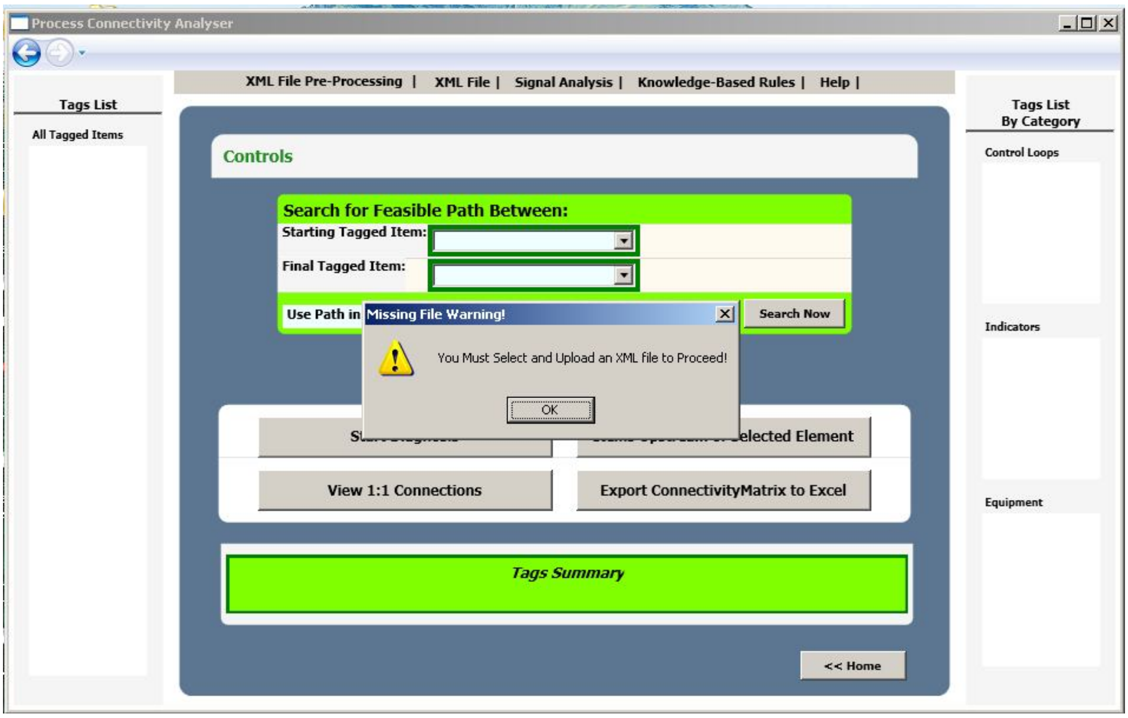


Figure 61: Error reporting for an invalid operation

In the subsequent screenshots, the results displayed are those of the illustrative example of Figure 77 treated in detail in Section 5.4. Figure 62 displays the error checking mechanism that prevents the user from choosing the same tag as the start and end tag.

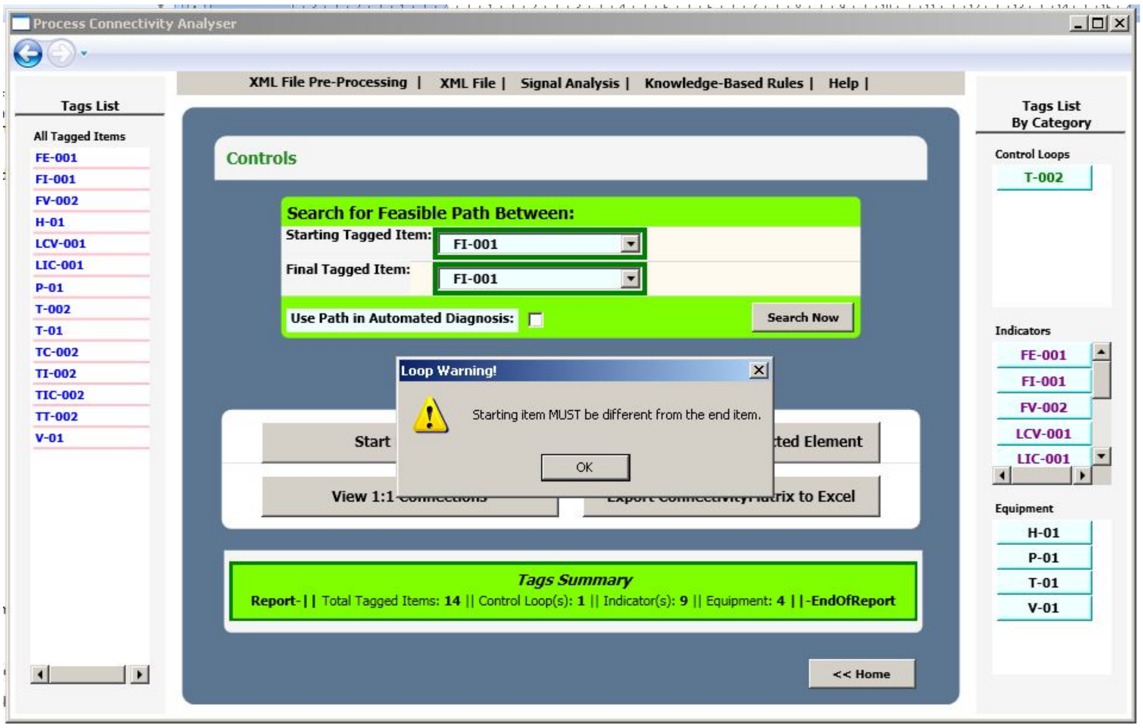


Figure 62: Integrity check to ensure that the starting tag is different from the end tag

Figure 63 presents a sample search results with Use Path in Automated Diagnosis: ☑ checked. Here, no path was found between the chosen tags Figure 63 shows a path search result with two paths found.

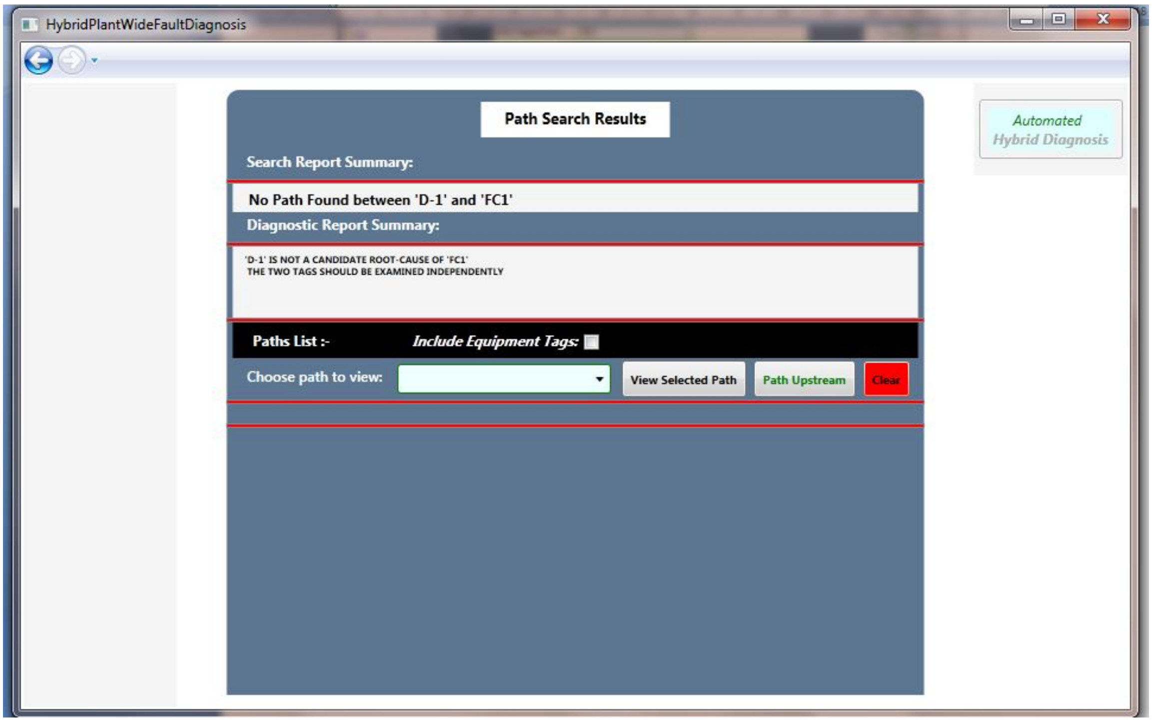


Figure 63: Sample path search result with diagnostic display

The user has a number of options here. Each individual path can viewed. The user has the option to include equipment the result display. The default display list the instrument (indicators and controllers). The user can further choose an item in the search list to see tags upstream of such chosen tag as shown in Figure 64.

Figure 64: **Path search display**

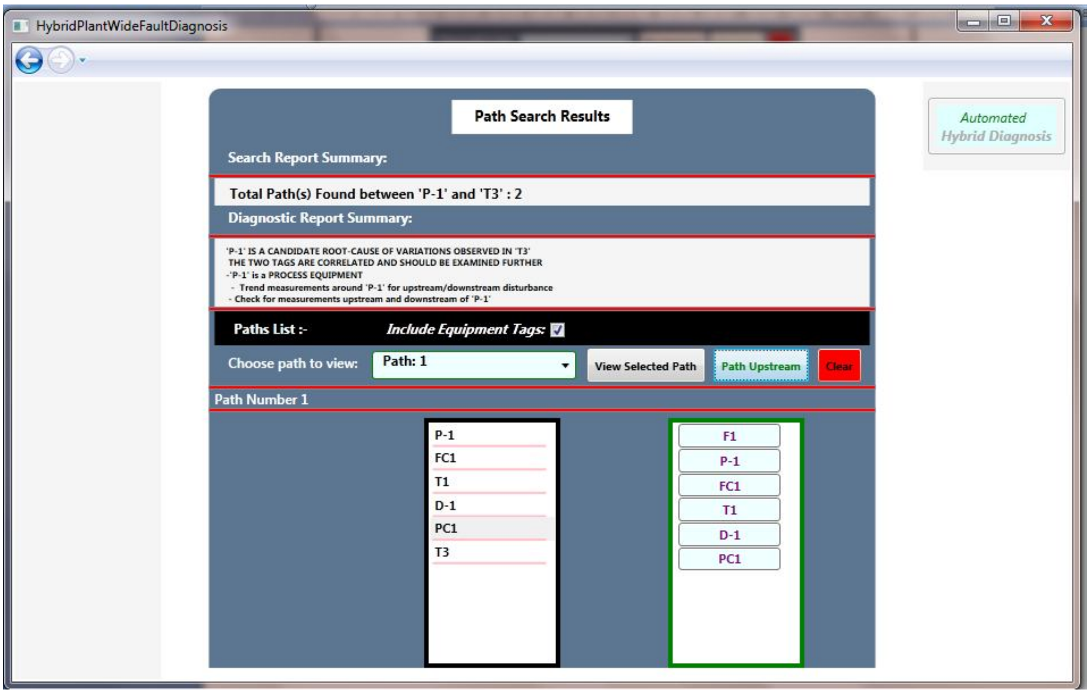
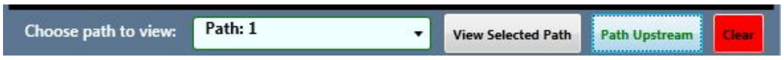


Figure 65: Path list display and upstream tags display

The clear button on the panel clears the upstream display.



Hybrid Diagnosis

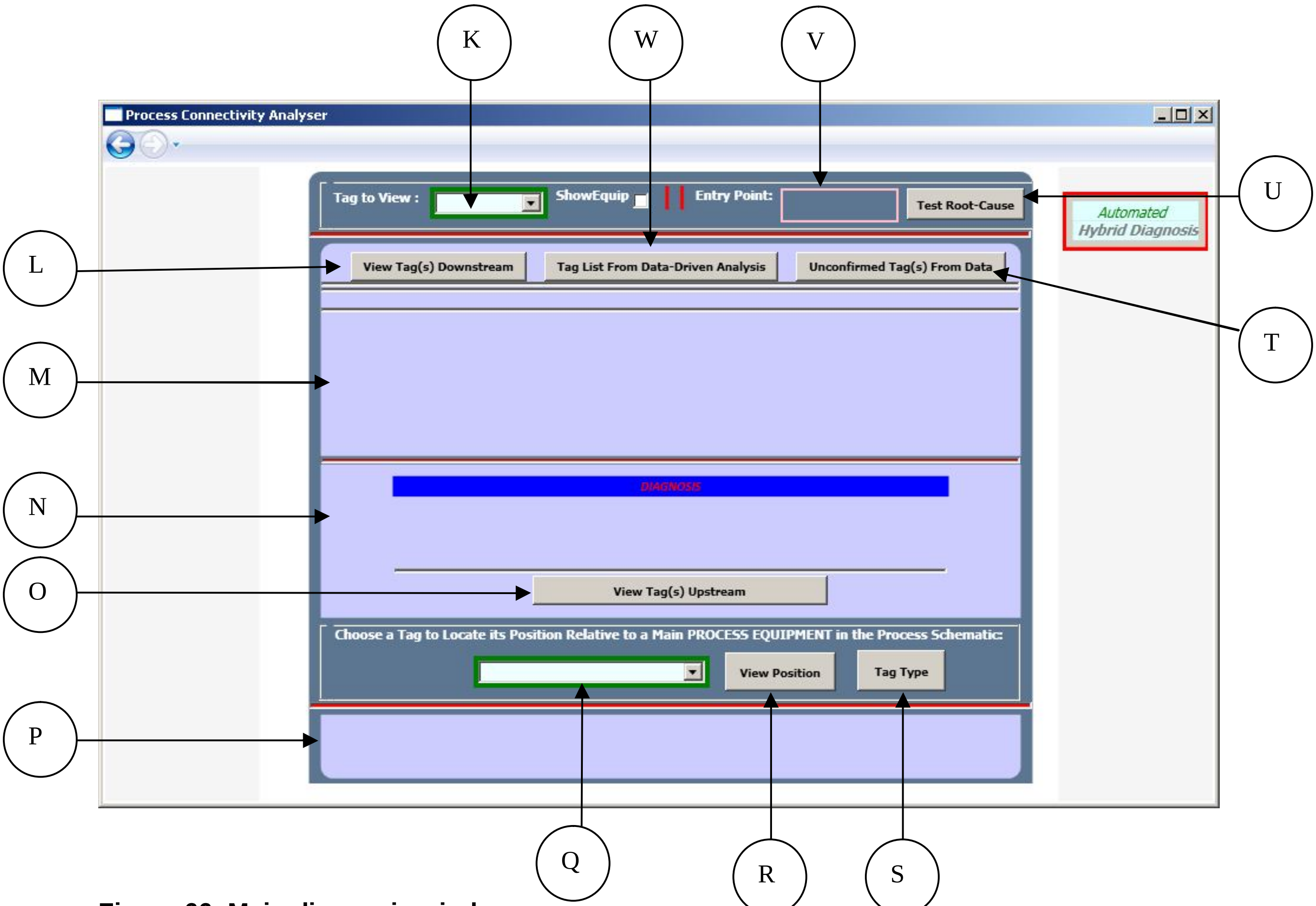


Figure 66: Main diagnosis window

(K) Choosing a tag from this drop down list allows the user to display downstream tags from the chosen tag by clicking on button (L). The results are displayed in (M)

(L) This button allows the user to display tags downstream of the chosen tag in (K). The system will prevent the user from performing this operation without selecting a tag from the drop down list first.

(M) This is the display area for either (L) (T) (U) or (W) button.

(N) This area displays diagnostic results including "Candidate Root-Cause" of distributed, plant-wide diagnosis.

(O) This button is for viewing tag items upstream of a chosen plant item in (K). A new window opens to display the result. Check is in place to prevent this operation without first choosing an item.

(P) Display area for results from operations in (Q) (R) and (S)

(Q) This user interface allows user to choose a tag from drop-down list to either check the tags position relative to a main process equipment such as a tower or vessel or to check whether the chosen tag is a controller, indicator or equipment. The results are displayed together or individually in (P)

(R) This functionality allows the user to view the relative position of a chosen tag from the list in (Q) to process equipment

(S) The button allows the user to view the type of the item selected in (Q) as an indicator, control loop or equipment.

(T) This operation displays tags in data analysis results that do not fall on the physical connectivity path.

U Hybrid fault diagnosis functionality-combination of quantitative process measurement analysis, qualitative connectivity information and process know-how is carried out by clicking this button. The result obtained will depend on the user's intent. It can be used to test hypothesis from data-based root-cause diagnosis, in which case, the user does not need to enter any input in V. The tool report the percentage of tags in data-driven analysis found on physical propagation path. Another use of this functionality is to find a candidate root-cause among a cluster of disturbed tags. In this case, the user will need to enter the entry tag. The entry tag is the inlet measurement point of process fluid (feed) or utilities such as steam, into the process. If there are multiple entry points, each must be tested to determine which one locates the highest percentage of disturbed tags on the feasible propagation path. The candidate root-cause of the report by the tool that gives highest percentage of confirmation will be the root-cause of such disturbed cluster of tags.

V This allows the user to enter an entry point as explained above. The text field is case sensitive which means that the entry must exactly match the starting tag the user intent to use.

W This functionality avails the user the opportunity to see a display of data-driven file loaded earlier in the main control window.

An attempt to perform root-cause diagnosis by clicking on U without first loading the data-driven result data file results in error message shown in Figure 67.

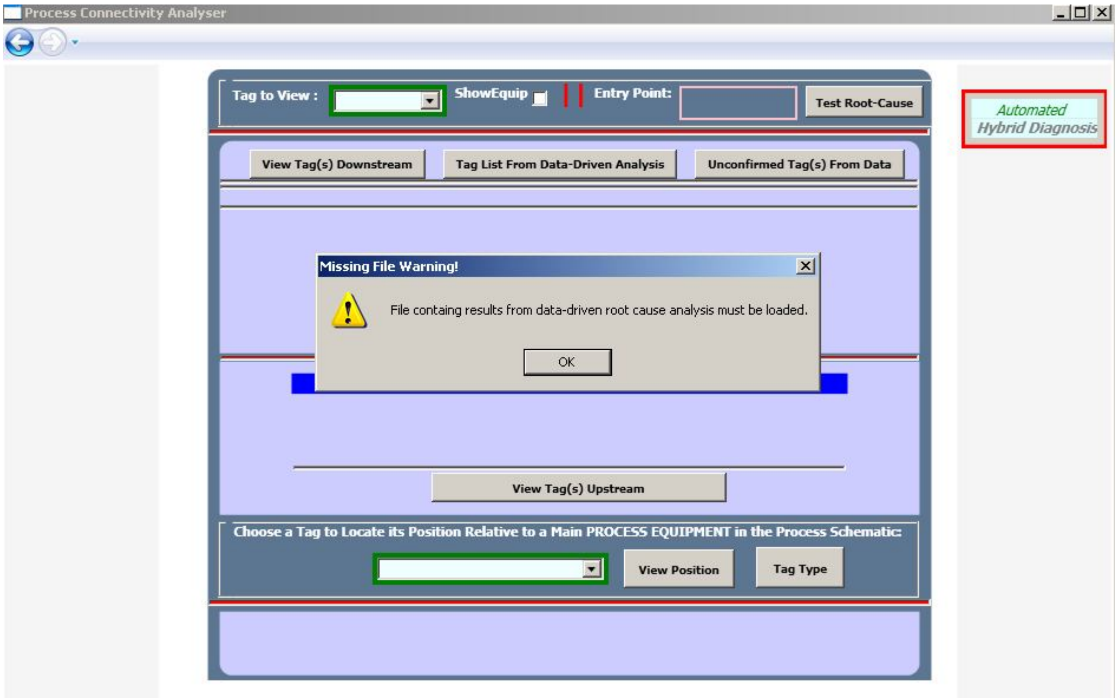


Figure 67: Error report for missing data file

Figure 68 shows a sample diagnosis result. The tool used process schematic XML format and data-driven cluster analysis to draw a conclusion.

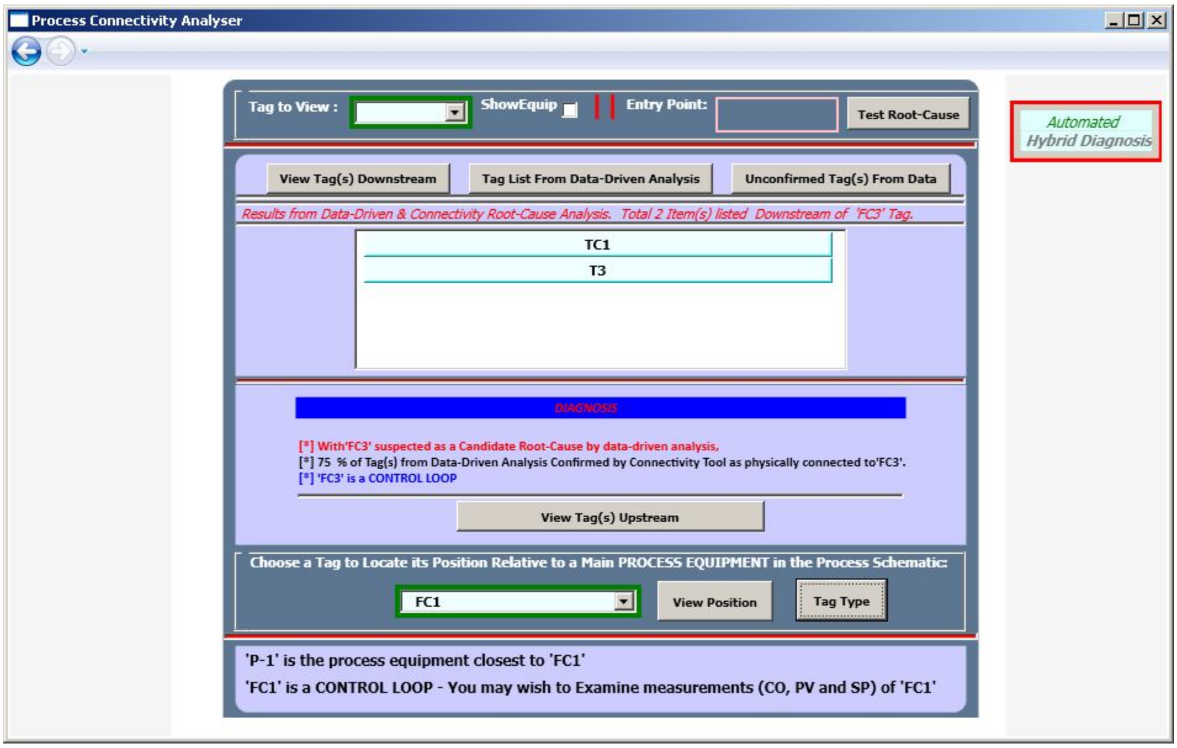


Figure 68: Hybrid fault diagnosis result

5.3.2 Other Operations

As mentioned earlier, the tool can be used for a variety of operations besides fault diagnosis. For these operations, no data-driven analysis is required. The application only uses the connectivity information as contained in XML file loaded into the tool. This section describes such extra operations.

Finding Tags Upstream of a Chosen Tag on the Main Control Window

This functionality allows the user to view items upstream of a chosen tag on the main control window without going any further. This is important because if the user already knows or has a hypothesis about a candidate cause of a co-ordinated, distributed and plant-wide disturbance, the user might simply be interested in viewing plant items upstream of such candidate cause especially when the suspected candidate cause is a measurement point such as an indicator. Any tag, as shown in figure 11, can be chosen from the left hand list box of the main control window to check items upstream of such chosen tag. In Figure 69, T3 is the chosen tag.

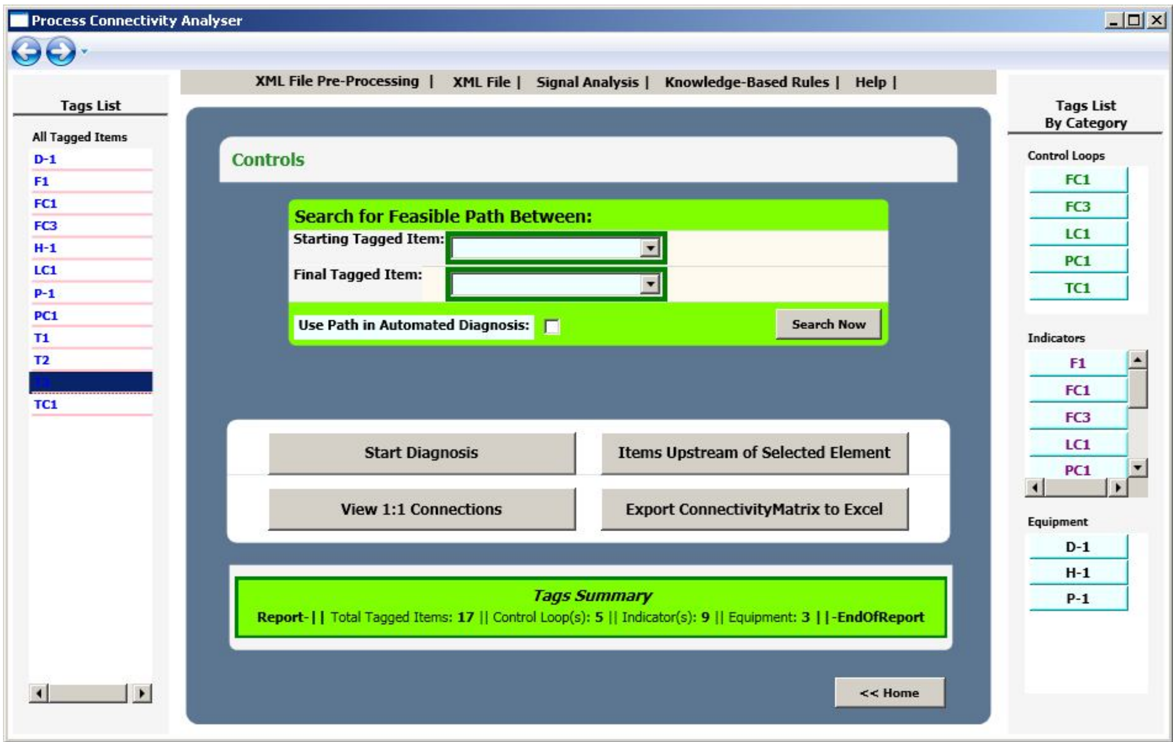


Figure 69: Selected tag to view the tag's upstream items

Once a tag is chosen, clicking on 

button produces the result shown in Figure 70

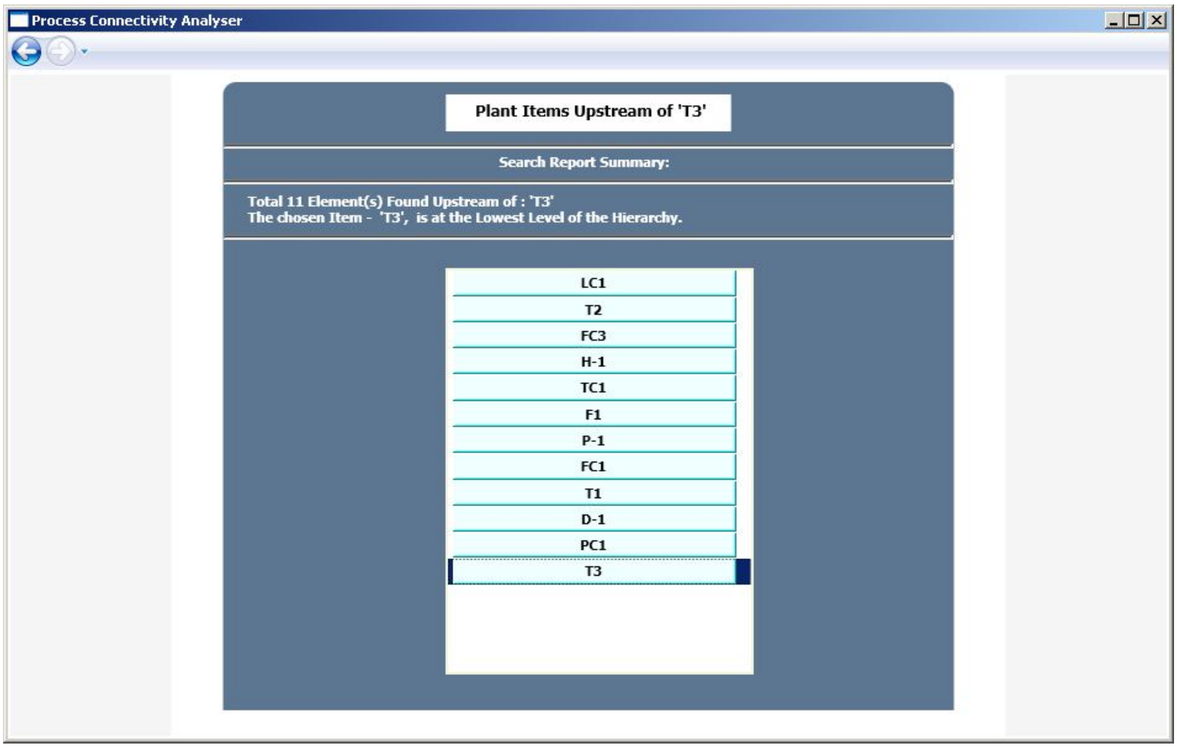


Figure 70: Upstream tags display window

<u>Checking Full Connectivity at P&ID Drawing Stage</u>

The ability to check full connectivity is very useful during the process of creating the intelligent P&ID to ensure that there are no loose ends. Drawings that are not fully connected are very difficult to visually detect on the drawing. When such a drawing is exported, the XML will contain insufficient connectivity information and will always produce spurious results. The tool allows this to be checked through the “View 1:1 Connections” 

functionality on the main control window. An example of a fully connected drawing is shown in Figure 71.

When a loose end is detected, the loose end is replaced with *UnidentifiedItem* to alert the user of a loose end and recheck the drawing again. In Figure 72, some loose ends were discovered and flagged as *UnidentifiedItem*. Two of such loose ends are circled in the Figure 72.

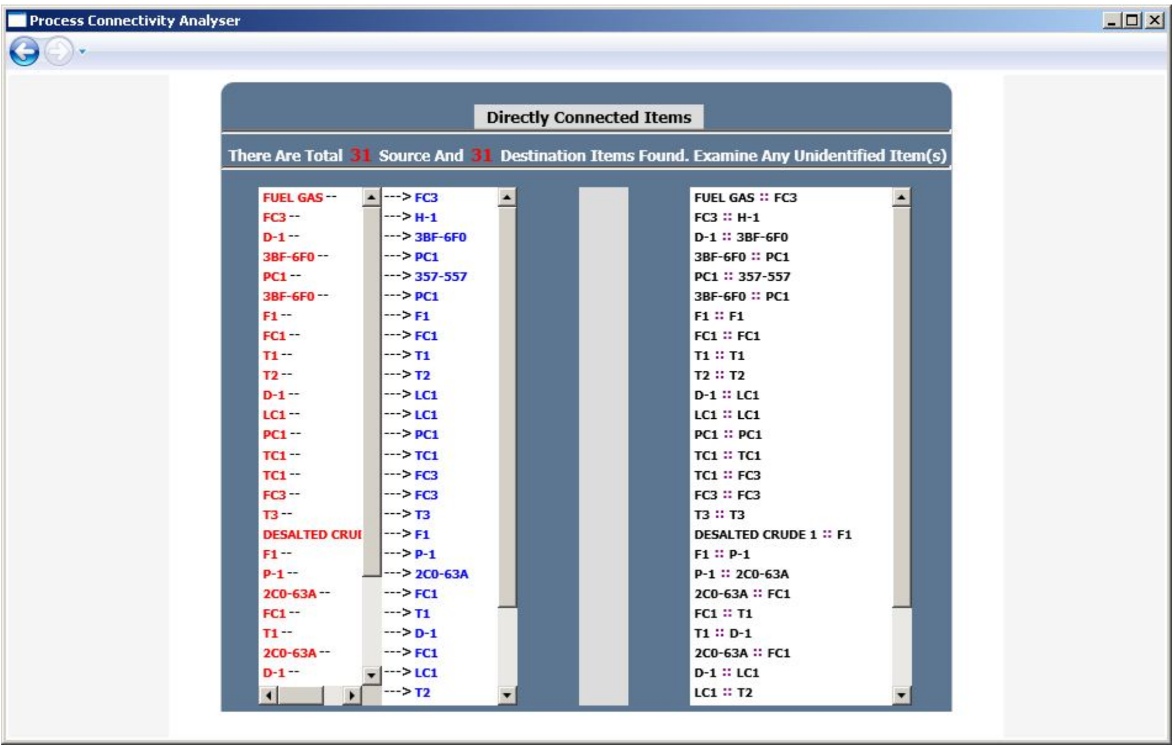


Figure 71: Checking for a loose end in the XML export

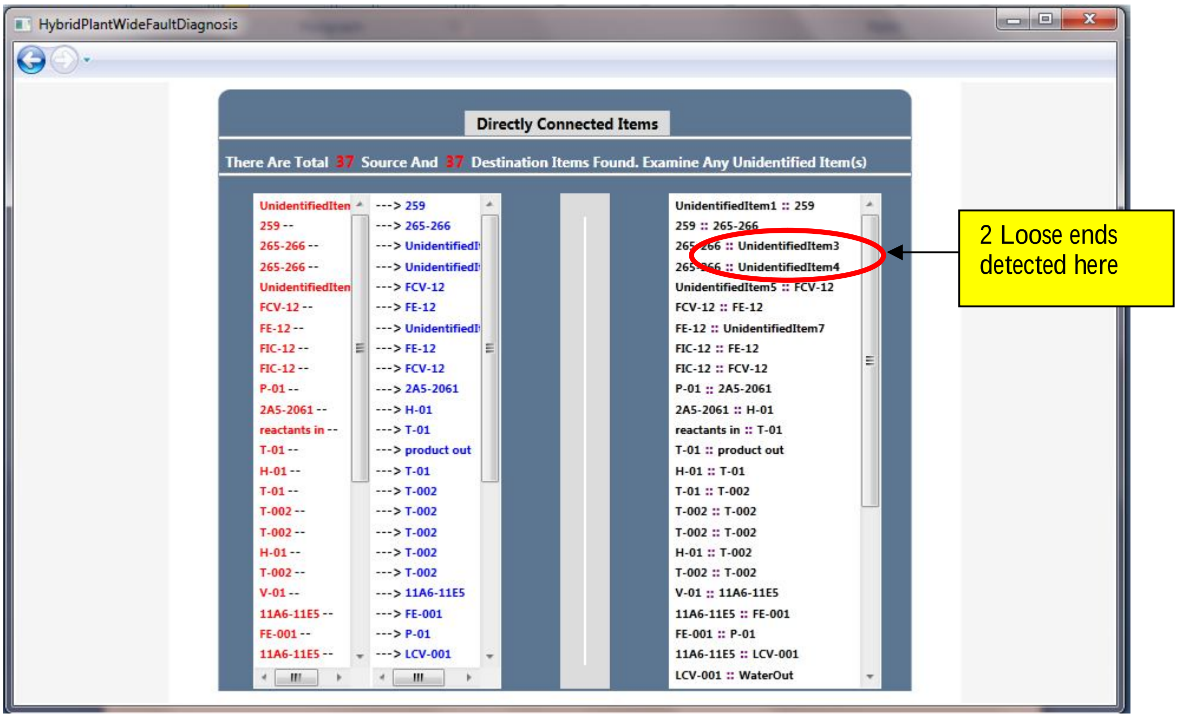
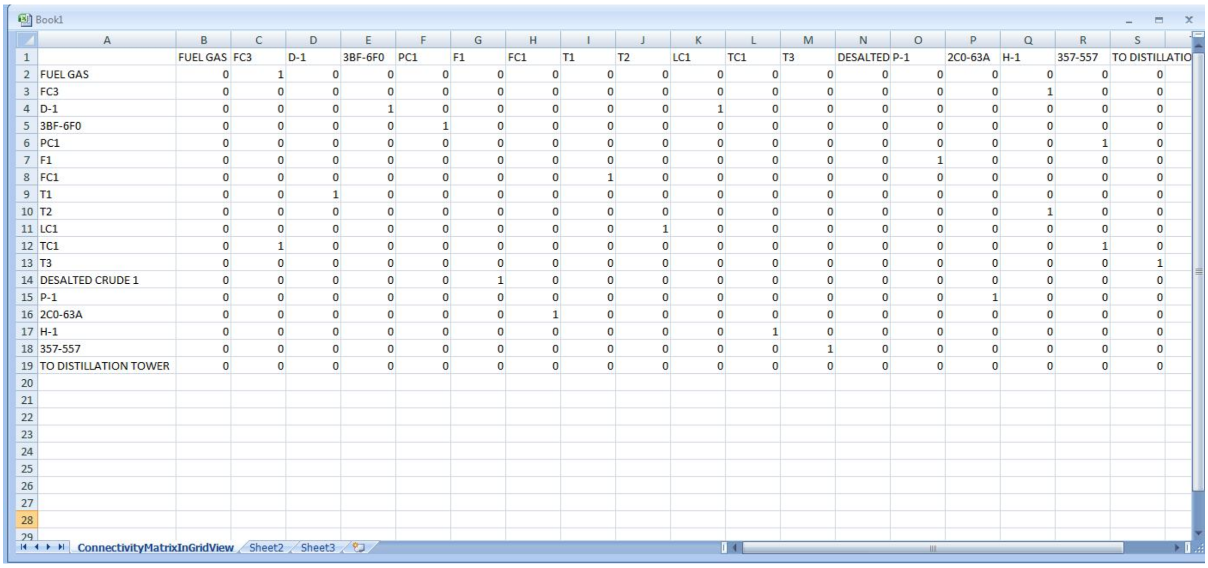


Figure 72: Process schematic P&ID drawing export with some loose ends detected as unidentified items

Connectivity Matrix in Excel Application

One of the intermediate results of the application that can serve as a basis for other application is the connectivity matrix of the P&ID, with “0” at intersection indicating no connection between the element on the row and that on the column of the intersection. The connectivity matrix can be exported to Microsoft Excel application as shown in Figure 73.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		FUEL GAS	FC3	D-1	3BF-6F0	PC1	F1	FC1	T1	T2	LC1	TC1	T3	DESALTED	P-1	2C0-63A	H-1	357-557	TO DISTILLATIO
2	FUEL GAS	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	FC3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
4	D-1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	3BF-6F0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
6	PC1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
7	F1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
8	FC1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
9	T1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	T2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
11	LC1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
12	TC1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
13	T3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	DESALTED CRUDE 1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
15	P-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
16	2C0-63A	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
17	H-1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
18	357-557	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
19	TO DISTILLATION TOWER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 73: Connectivity matrix exported to Microsoft Excel Application

5.3.3 Operations on Multiple P&IDs and Modular Approach

Another useful functionality provided by *Process Connectivity Analyser* is the ability to combine multiple P&IDs for analysis. This user’s option is circled red in Figure 74. Conversely, a section of a large P&ID can also be isolated for analysis thus providing great modular flexibility.

Figure 74: Functionality merging multiple P&ID drawing files

Here modularity implies that a large and complex plant can be broken down for analysis, for example as drawn on separate P&IDs, for a more focused analysis.

This concept of file merging via off sheet connectors is depicted in Figure 75.

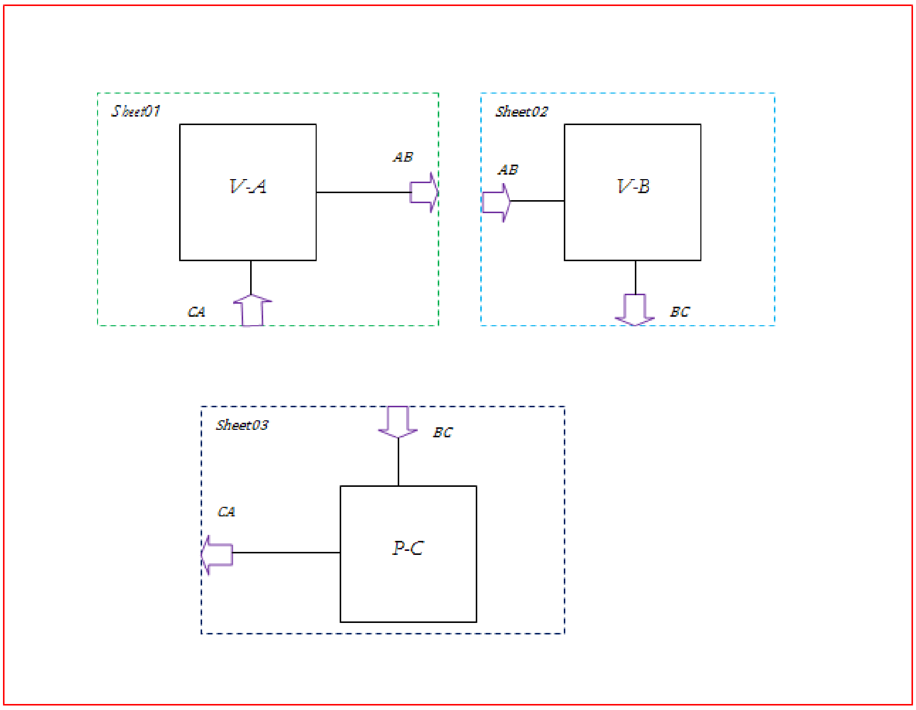


Figure 75: Schematic drawing to depict P&IDs merging and modularity

Three sheets of process plant P&IDs are shown in Figure 75. Each individual drawing can be exported separated as XML. The *XML File Pre-Processing* tab combines all the three files together to generate a single XML file for analysis.

5.4 Illustrative Example-Simplified Atmospheric Crude Heating Unit

This section presents an illustrative example to demonstrate and prove the concepts described in the thesis. The chosen example is a simple crude heating unit to aid easy understanding of the material. More complicated real life example follows the relatively simplified illustrative example discussed in this section.

5.4.1 Process description

The example presented in Figure 76 is a process P&ID of a crude heating unit of an atmospheric crude distillation schematic shown in Figure 78. The process P&ID is drawn with intelligent CAD (AVEVA) running on AutoCAD software. The same process schematic is reproduced in Figure 77 for clarity. The example, circled in Figure 78, is chosen to demonstrate automated cause-and-effect relationships among plant items and shows that this can quickly become complicated even for such a simple process. The concept of downstream and upstream with physical propagation paths for process fluid and control and measurement signals will also be investigated.

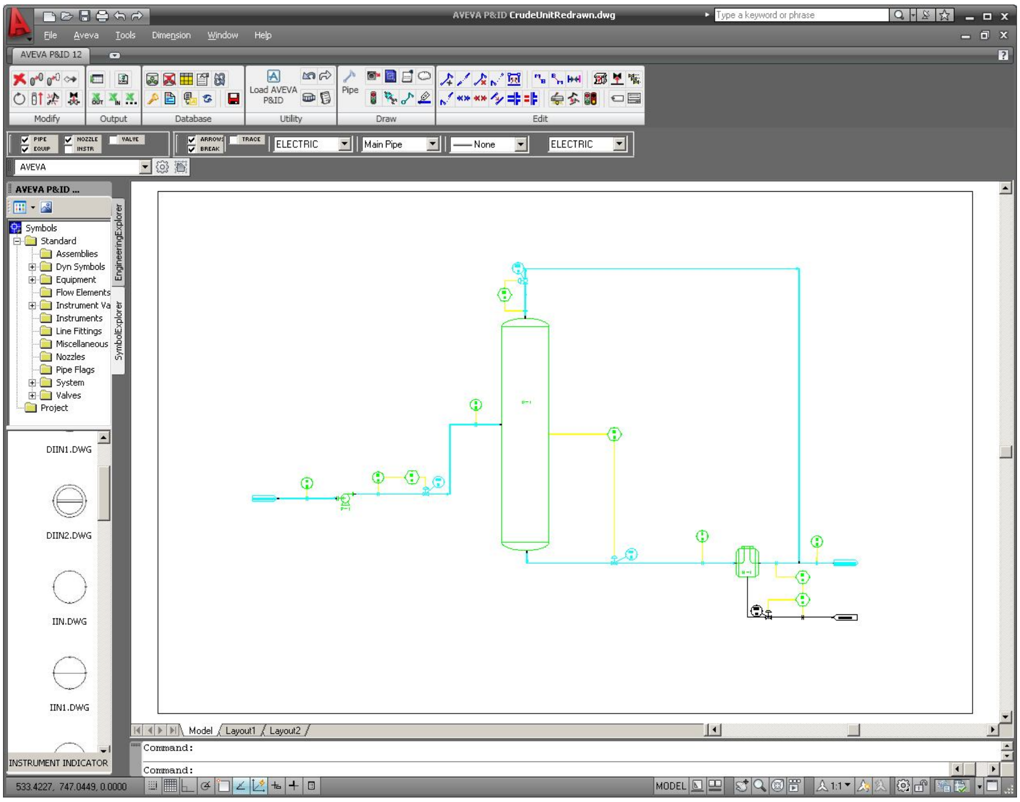


Figure 76: Crude heating unit process P&ID created with AVEVA intelligent CAD tool running on AutoCAD

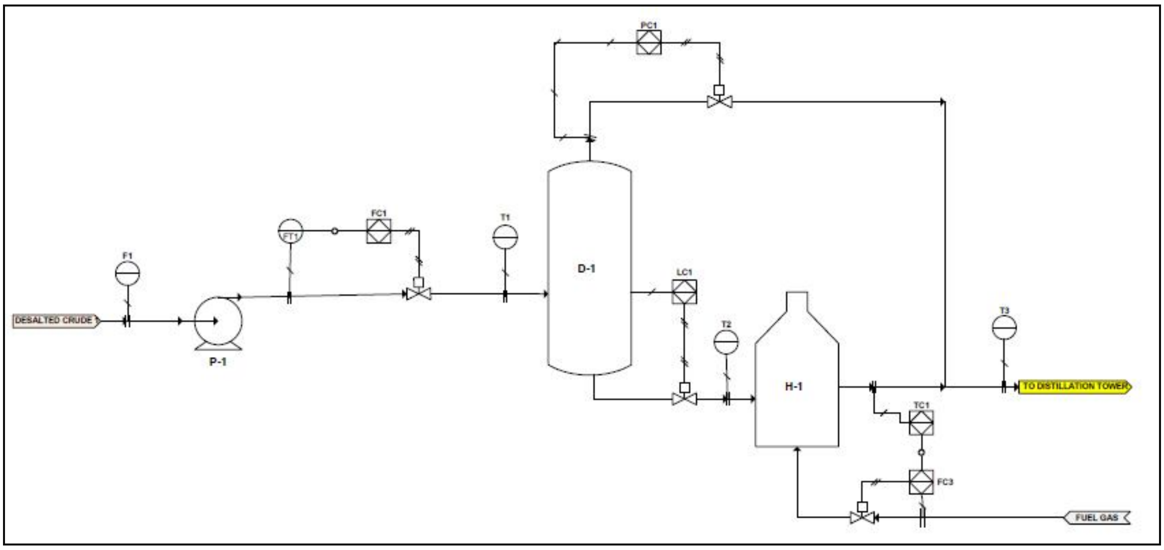


Figure 77: Process schematic with instrumentation of a simplified crude heating unit

For the heating unit process, desalted crude is pumped by the charge pump P-1 into a pre-flash drum D-1. Flashed crude from the bottom of D-1 enters the feed pre-heat furnace H-1 and the vaporized mixture is sent to the feed tray of the atmospheric tower of the distillation unit. Temperature of the heated crude feed to the atmospheric distillation tower is controlled by regulating the flow of fuel gas supply to H-1 through a cascade control system with temperature controller TC1 as the master controller dictating set-point values to flow controller FC3, the slave controller in the cascade loop. Level control loop LC1 controls the level of process fluid in D-1 while pressure control loop PC1 controls the pressure in D-1. Desalted crude flow into D-1 is controlled by flow control loop FC1. F1, T1, T2 and T3 are indicators.

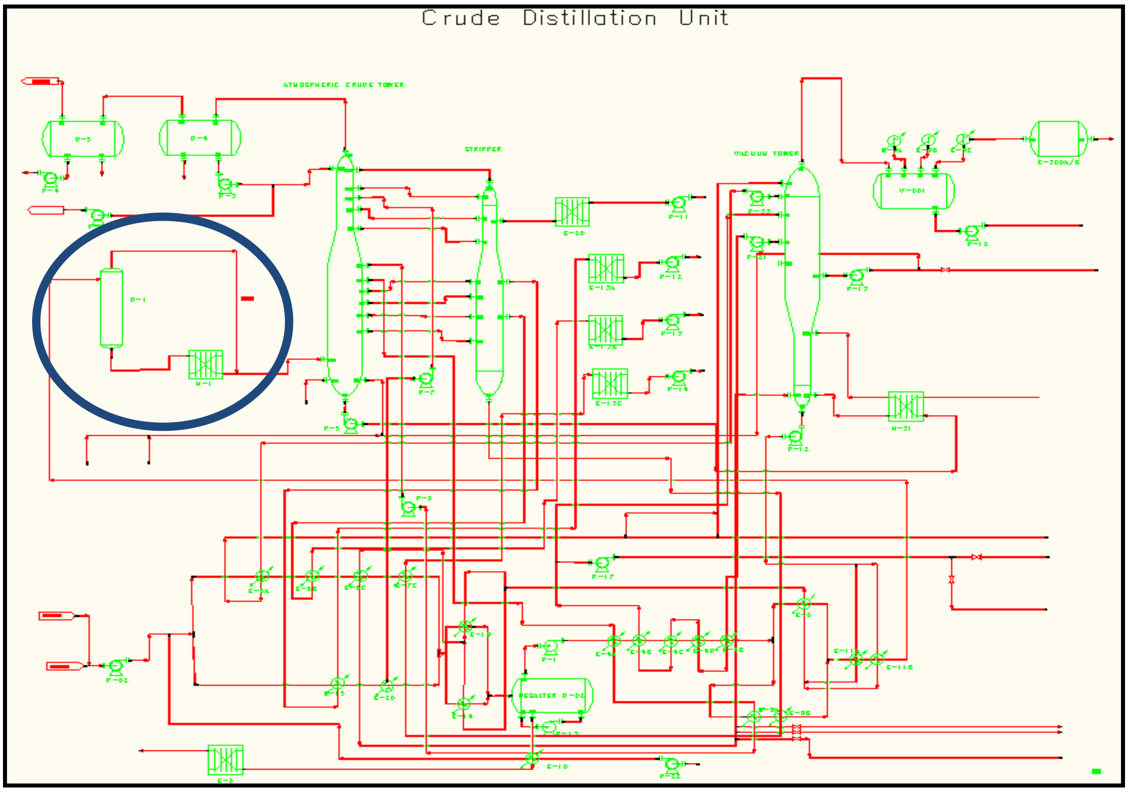


Figure 78: Crude distillation schematic with crude heating section, the illustrative example, of the unit encircled

The following sub-sections describe the sequence of steps taken to carry out analysis on the atmospheric crude heating unit shown in Figure 77.

5.4.2 XML description of the process schematic

In order to automate the analysis, the intelligent process schematic was converted to electronic text-based XML representation. AVEVA P&ID CAD tool has functionality that allows for ISO15926 compliant XML export of intelligent process P&IDs. Part of the XML representation is shown in Figure 79 in stylus Stylus Studio® 2007.

The XML file contains all the drawn plant items, the connections among the plant elements and other information that are not relevant to the current work

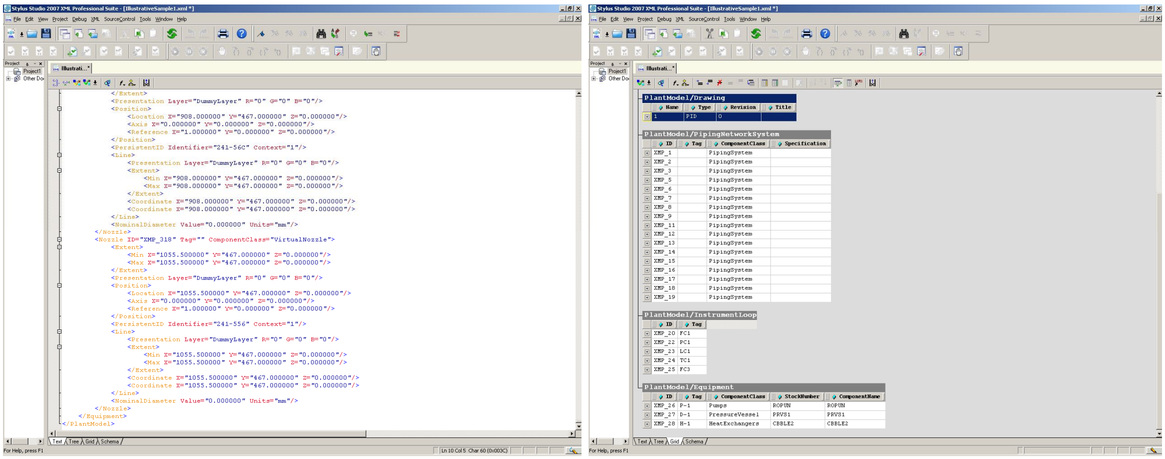


Figure 79: Text-based XML representation of the simplified crude heating unit. Text view shown on the left hand side and grid view on the right hand side in Stylus Studio® 2007.

The text file was loaded and processed by the parser, *Process Connectivity Analyser,* to automatically generate the process connectivity matrix, a form of computer representation of directed graph shown in Figure 80. The connectivity matrix can be used an independent module and starting point for other application (Alabi, 2010) and (Di Geronimo Gil, 2010).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		FUEL GAS	FC3	D-1	3BF-6F0	PC1	F1	FC1	T1	T2	LC1	TC1	T3	DESALTED CRUDE 1	P-1	2C0-63A	H-1	357-557	TO DISTILLATION TOWER
2	FUEL GAS	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	FC3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
4	D-1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	3BF-6F0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
6	PC1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
7	F1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
8	FC1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
9	T1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	T2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
11	LC1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
12	TC1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
13	T3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	DESALTED CRUDE 1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
15	P-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
16	2C0-63A	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
17	H-1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
18	357-557	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
19	TO DISTILLATION TOWER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 80: Automated connectivity matrix of the crude heating example, exported to Microsoft® Excel ® application

5.4.3 Signal Analysis

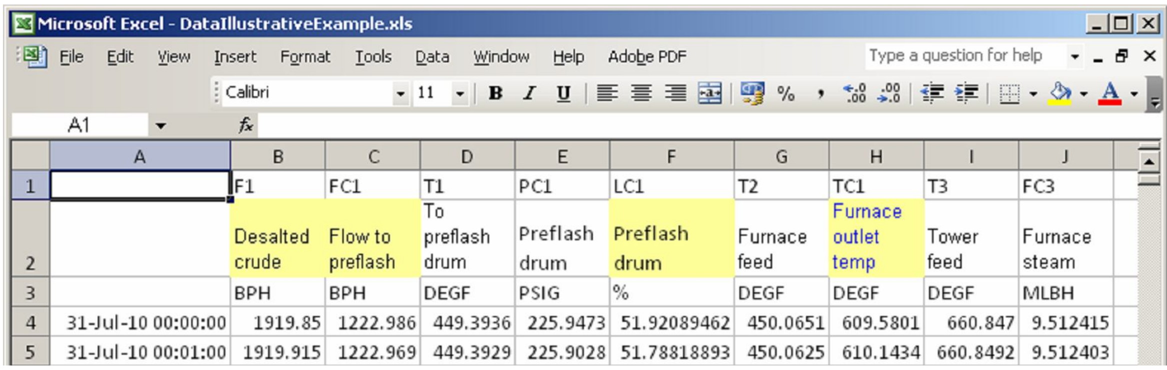
This section describes the PDA data analysis for the atmospheric crude heating unit process described in Figure 77. The PDA analysis described in this section is detailed because subsequent data analysis for the industrial case study in section 5.5 undergoes similar treatment without the need to repeat every step discussed here.

As mentioned earlier, the PDA signal processing tool used for the data analysis is a commercial tool from ABB, a research output from the Imperial/UCL Centre for Process Systems Engineering (Bauer, *et al*., 2007; Bauer and Thornhill, 2008; Thornhill, 2005; Thornhill, *et al*., 2003; Thornhill, *et al*., 2002). The university has a PDA license in place for academic use.

The workflow is to run data-driven analysis and to test its hypotheses with the connectivity and directionality information from the *Process Connectivity Analyser.* The combined approach is an example of hybrid modelling and process analysis.

The Data Set

Plant data were sampled from the control system for each of the measurement points. Two days worth of data sampled every minute was used in the PDA data analysis. A fragment of the data suitable for PDA tool is shown in Figure 81. The data are in an Excel spreadsheet with time stamps in the first column and the measurements tags in subsequent columns. The numerical data values start in the fourth row and the first three rows are reserved for other information.



	A	B	C	D	E	F	G	H	I	J
1		F1	FC1	T1	PC1	LC1	T2	TC1	T3	FC3
2		Desalted crude	Flow to preflash	To preflash drum	Preflash drum	Preflash drum	Furnace feed	Furnace outlet temp	Tower feed	Furnace steam
3		BPH	BPH	DEGF	PSIG	%	DEGF	DEGF	DEGF	MLBH
4	31-Jul-10 00:00:00	1919.85	1222.986	449.3936	225.9473	51.92089462	450.0651	609.5801	660.847	9.512415
5	31-Jul-10 00:01:00	1919.915	1222.969	449.3929	225.9028	51.78818893	450.0625	610.1434	660.8492	9.512403

Figure 81: Data for crude heating unit example in PDA format

The data file in the format described above is loaded into the PDA tool for data-driven analysis. Figure 82 corresponds to a successfully loaded data with the details displayed on the PDA tool.

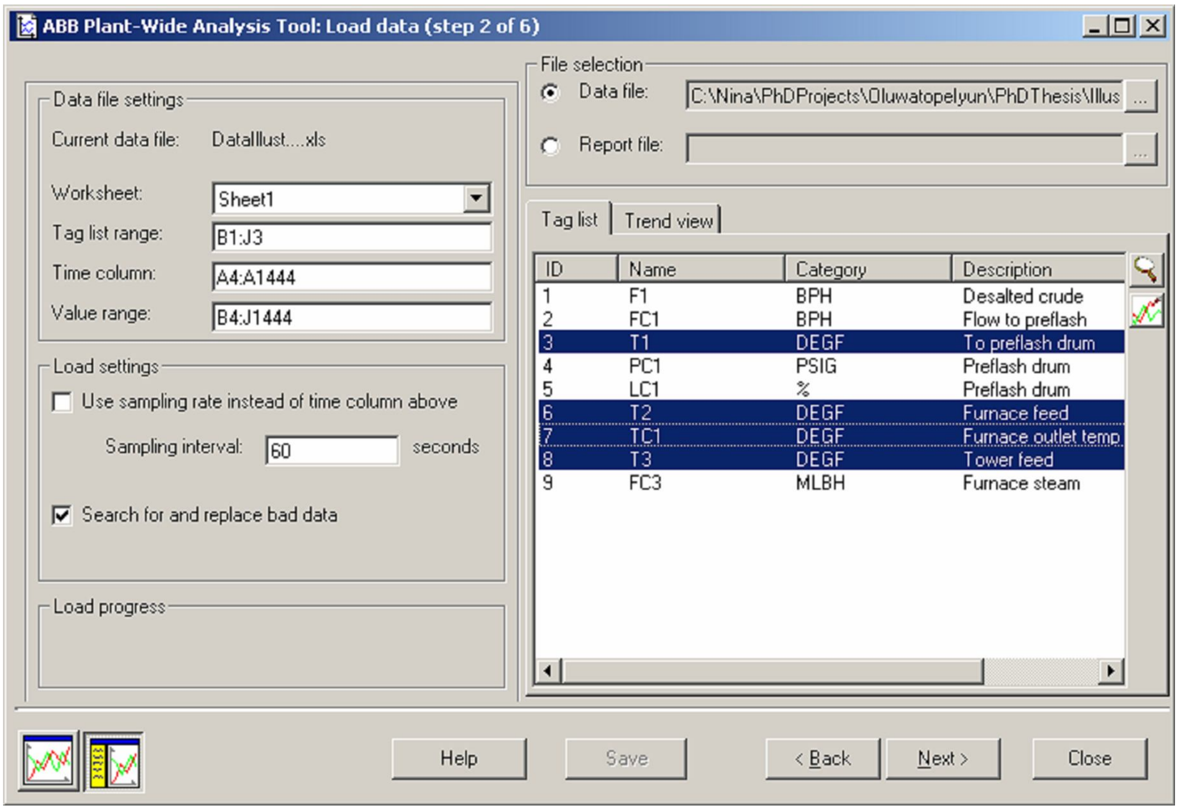


Figure 82: Data loaded into PDA tool with a description of tags

The data can be plotted and visualized as time trend as shown in Figure 83. This plot shows the absolute values; whereas subsequent plots show mean centred data scaled to unit standard deviation

Visual examination of the plot shows that temperature TC1 is lower than temperature T3, while T1 and T2 are similar and lower. These observations can be tested against the topology to see if they make sense. Hence, knowledge about process topology can be utilized to check for correctness before proceeding further in the data analysis.

The temperature is expected to be rising as desalted crude moves from the desalter through the pre-flash drum, D-1and the furnace, H-1 and ultimately to the atmospheric crude distillation tower.

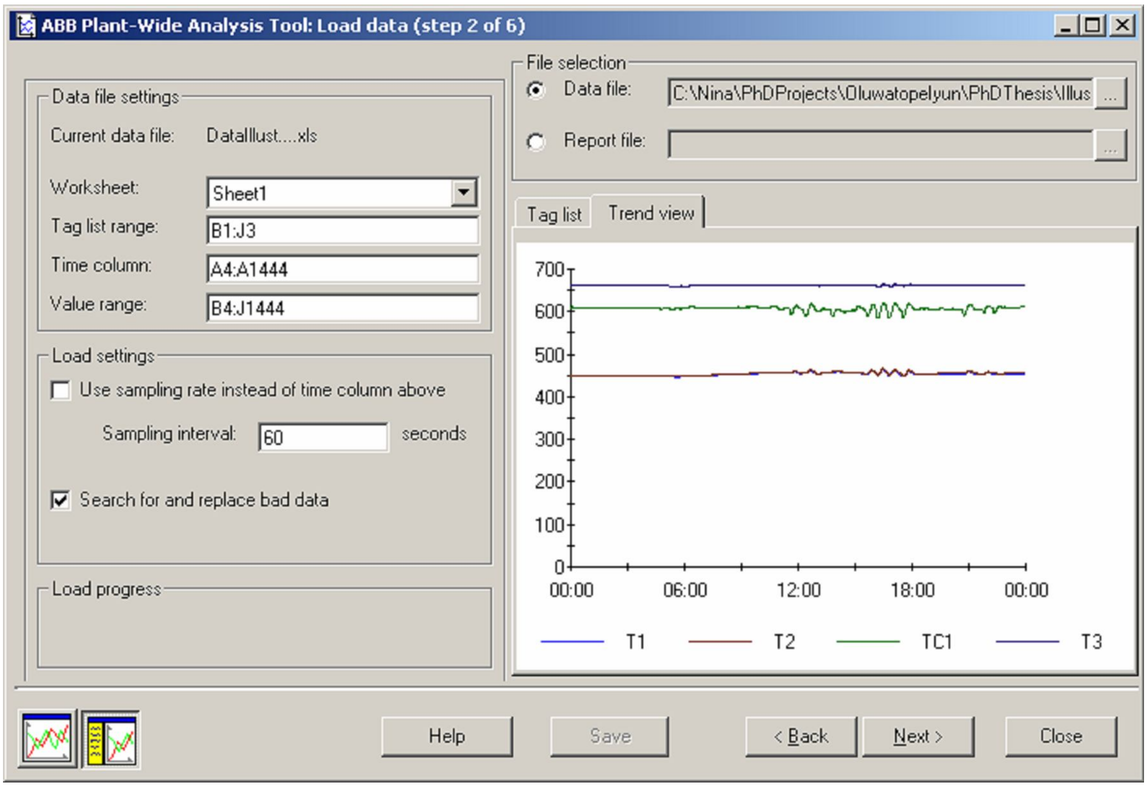


Figure 83: Data visualization in PDA tool. Absolute value time trend of selected process variables

Process connectivity information and know-how should indicate that temperature measurement of the crude feed to the atmospheric distillation tower, T3 should be the highest temperature of the crude. This can be verified using the connectivity tool as follows:

By selecting T3 and examining all crude temperature measurement upstream of T3. T1, T2 and TC1 should all be upstream of T3 (higher in the hierarchy with lower temperature with respect to T3). This is shown in Figure 84with T1, T2 and TC1 circled in red.

Another observation from the time trend is that TC1 has quite large deviations compared to T3, T1 and T2. This is a potentially important observation because later plots hide this fact when showing mean centered and normalized data.

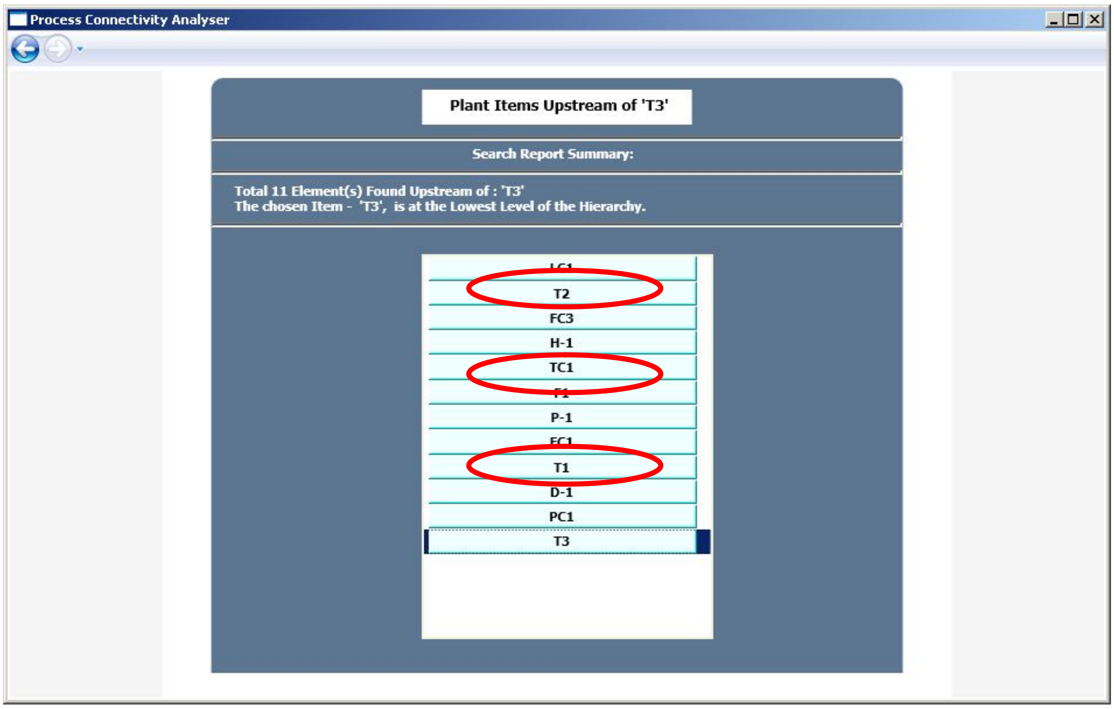


Figure 84: Temperature measurement points (in red circle) upstream of T3

Data Compression for Signal Analysis

The PDA tool has an inbuilt functionality to assesses whether the data are valid for analysis by calculating a compression factor. *Compression* refers to the technique of saving space in the data historian by only recording exceptional points, and then joining up the points with straight lines. If compression factor (CF) is >1 then data-driven analyses may not give the correct results. The analysis shows that some time trends are too compressed to be reliable.

As shown in Figure 85, F1, FC1, T1 and T2 all have CF >1 and have been excluded from the analysis. However, in the second plot depicted in Figure 86, compressed data have been included for the analysis because compression was not extreme and visual examination suggested the features in the data were real and not artefacts caused by compression.

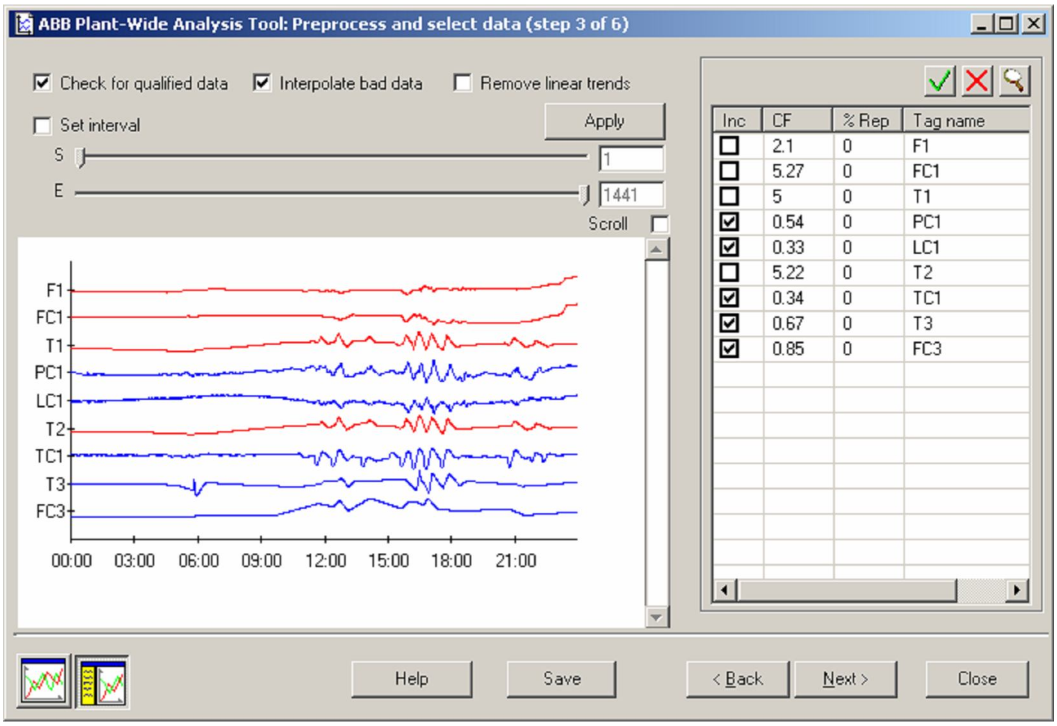


Figure 85: Time trend of uncompressed data measurements

Based on this observation and personal judgement, the compressed tags have been included in the analysis, even though the PDA recommended they should be excluded.

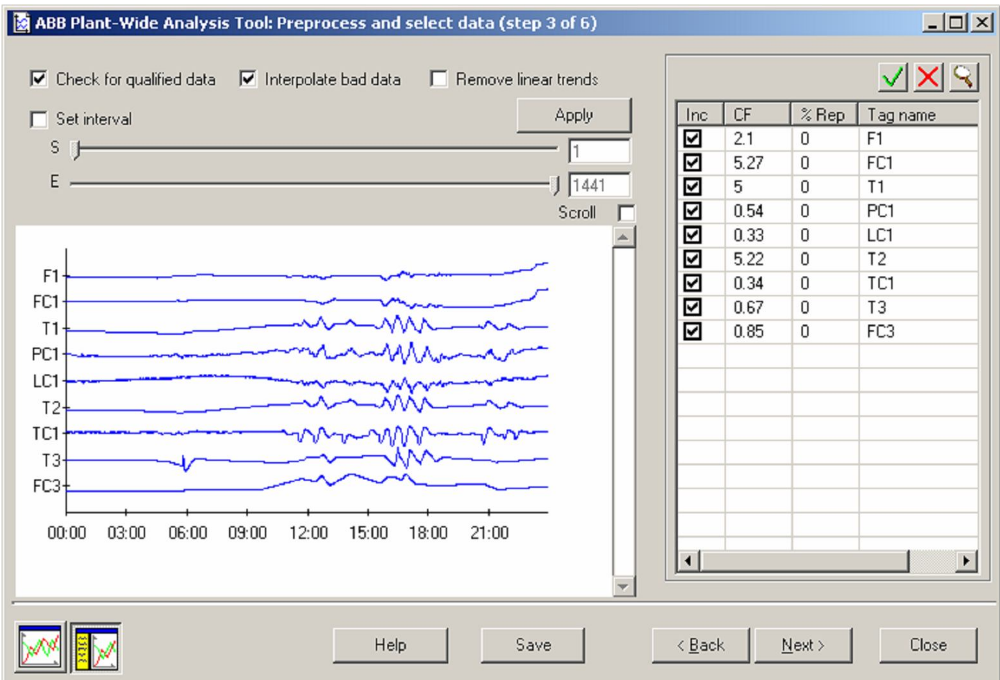


Figure 86: Time trend of both compressed and uncompressed data measurements

Analysis of Signal Frequency Content

Plots in Figure 87 show the time trends and also show their corresponding frequency spectra on the right hand side. The spectra on the right show frequency on the horizontal axis. The interpretation is as follows:

(i) any frequency content on the right hand side (e.g. towards 0.01 Hz) is noise in the measurement,

(ii) low frequency content e.g. towards 0.00001 Hz indicates that the time trend has an overall slope or curve. This can be seen prominently in the first two tags.

(iii) the frequency content in the middle ranges relates to the shorter term dynamic features in the data, for instance the oscillatory features occurring between 15:00h and 16:00h.

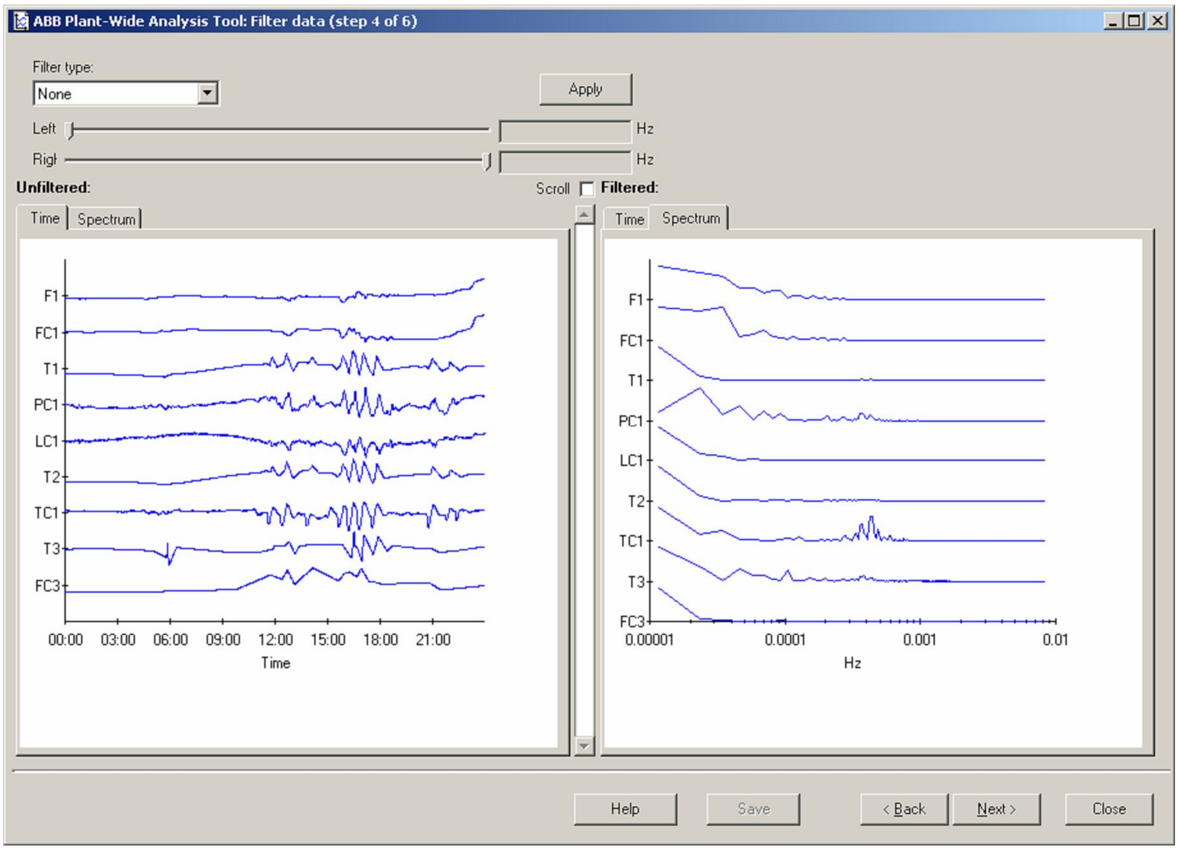


Figure 87: Plots of time trends and their corresponding frequency spectra

It is possible to examine the middle ranges more closely by filtering the data. A band pass filter removes the low frequency slowly-varying trends and also the high frequency noise. For the spectra analysis, filter boundaries have been set to 0.0001 and 0.001 Hz. On the right of Figure 88 are the filtered spectra which now have no high or low frequency content. The middle frequencies are therefore emphasised.

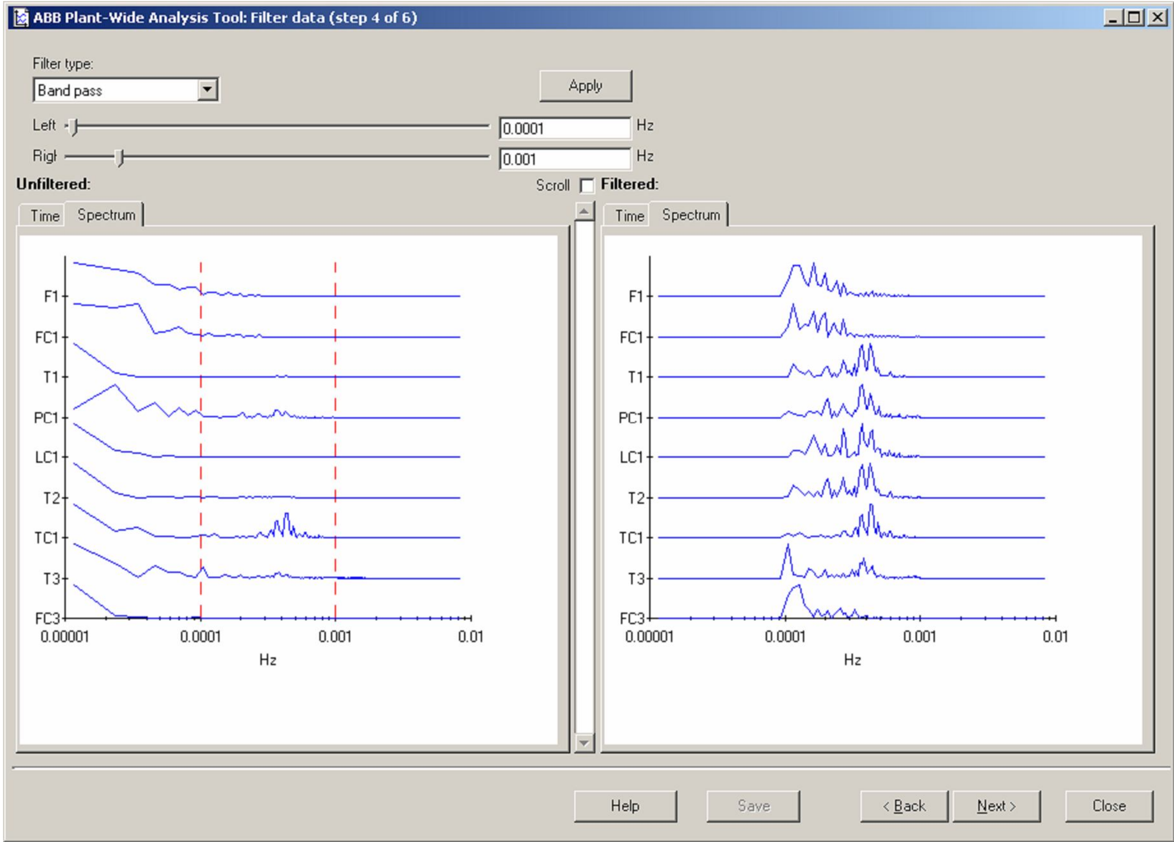


Figure 88: Plot of frequency spectra with band pass filter. The filter removes low frequency and high frequency noise components from the analysis

When viewed in the time domain, Figure 89 shows the time trends before and after filtering. The curves in tags F1 and FC1 have been eliminated to some extent, and the sharp triangular features in the compressed time trends have been smoothed.

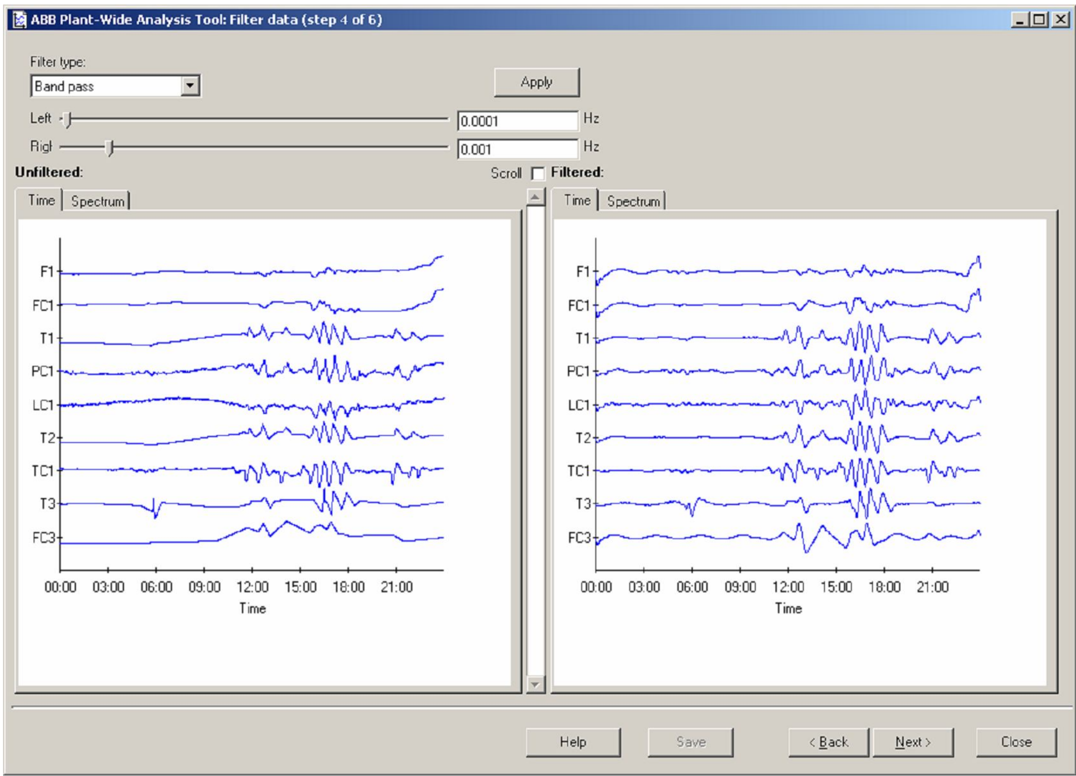


Figure 89: Time trends of filtered signals with compressed data included in the analysis

Root-cause Analysis

The following sections will first analyse the unfiltered data for the root-cause and then the filtered data. The results obtained will be compared with and validated against the process topology connectivity tool.

Root-cause analysis of unfiltered data

The clustering page shown in Figure 90 is using the unfiltered data. The PDA tool suggests that the time trends shown in red are similar to each other.

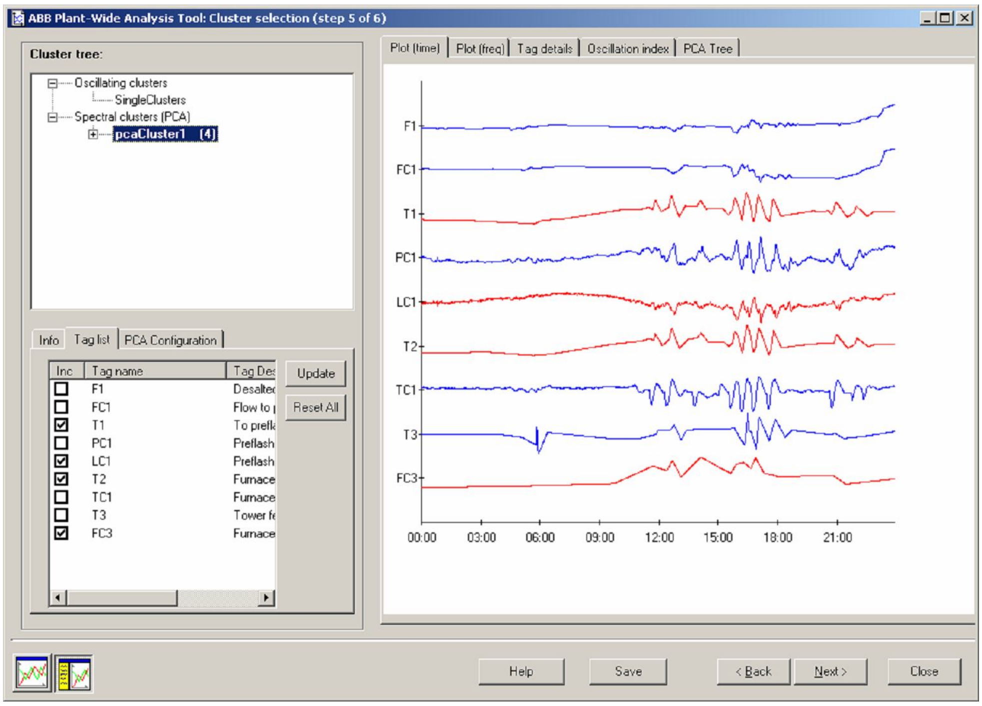


Figure 90: Clustering analysis for unfiltered data

However, visual inspection shows that some other tags also exhibit similar behaviour hence a custom cluster is created by selecting and updating additional tags in the lower left panel of the PDA tool page shown in Figure 91.

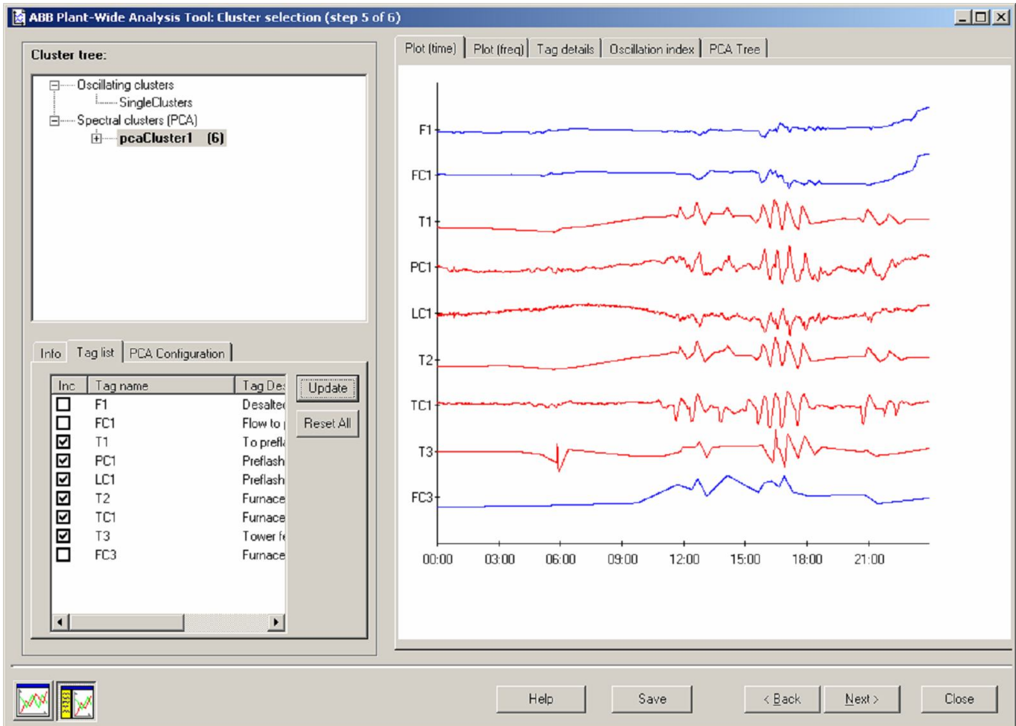


Figure 91: Custom cluster created to include tags with similar behaviour

Figure 92 shows the root-cause analysis obtained using the custom cluster.

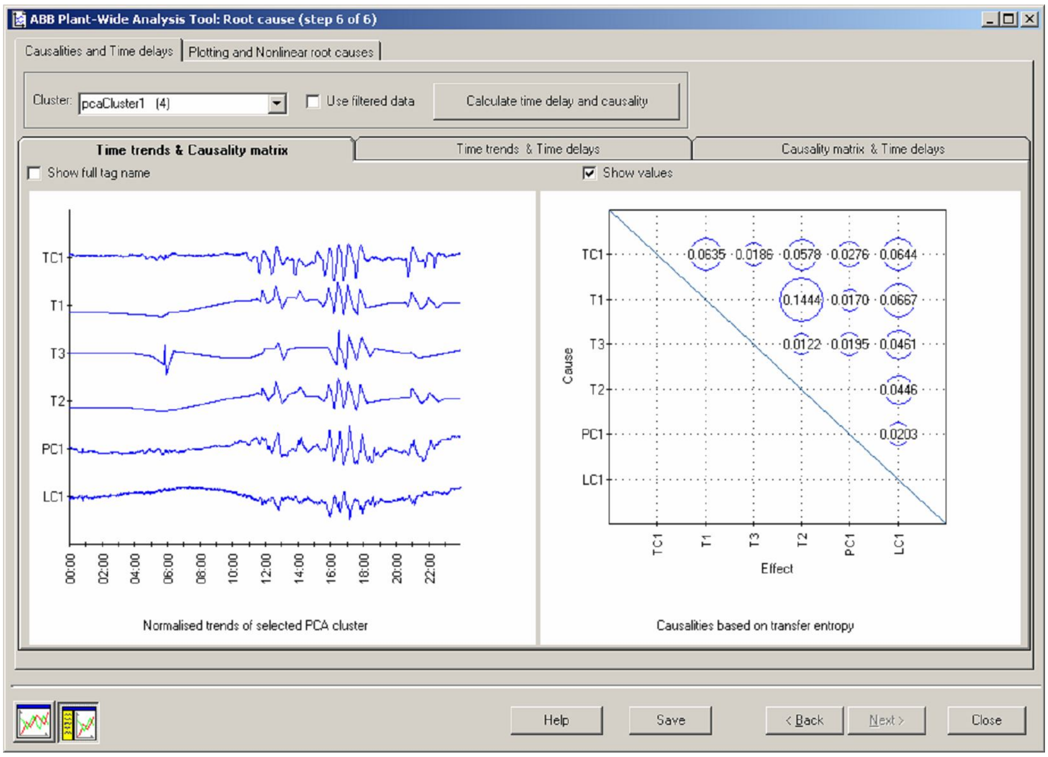


Figure 92: Unfiltered data-driven root-cause analysis for the atmospheric crude heating example.

The result suggests TC1 as the root-cause with T1, T3, T2, PC1 and LC1 as secondary propagated effects.

Testing data-driven root-cause hypothesis against the process topology and connectivity

The results suggested by the data-driven analysis tool need to be checked against the process topology to ensure that they make sense. Results from hybrid data-driven and causality analysis shown in Figure 93 indicates that of the five measurement points (T1, T3, T2, PC1 and LC1) where secondary propagated effect are observed according to data-driven analysis, only one tag (T3) is physically connected to the suggested root-cause (TC1) representing just 33.33% of tags where secondary propagated effects are observed.. Thus there is no physical propagation path from TC1 to either T1, T2, PC1or LC1 as suggested by data-driven analysis.

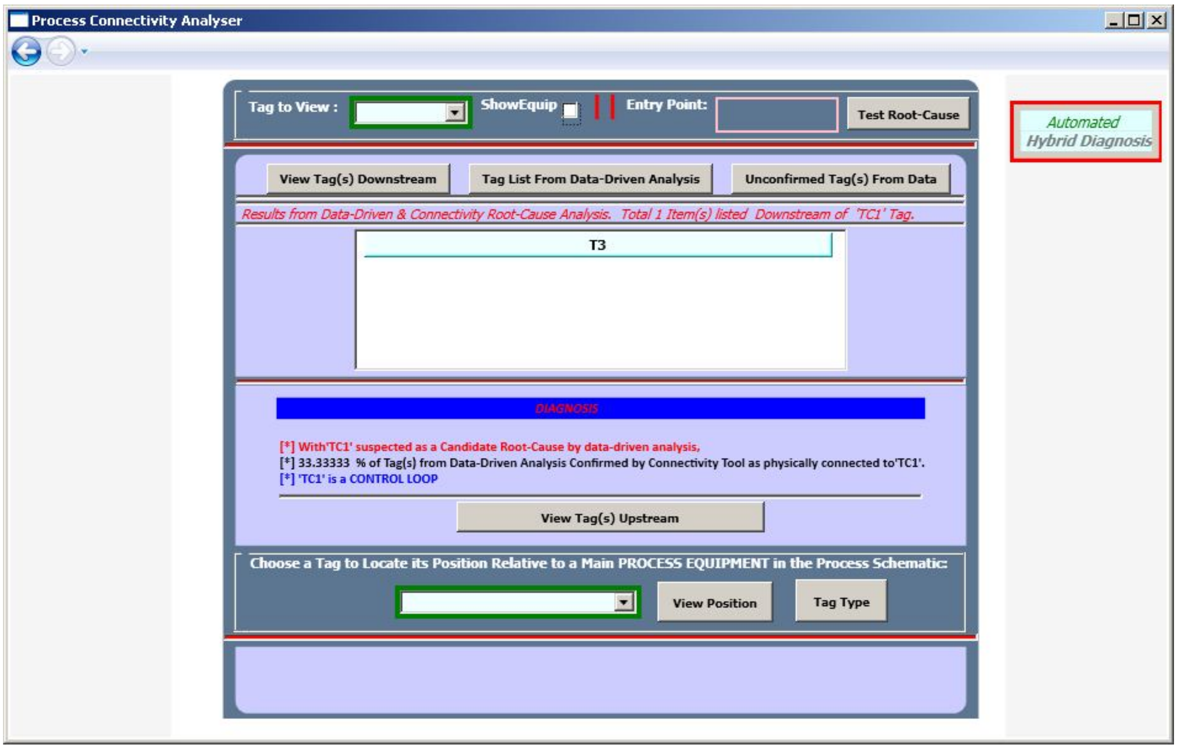


Figure 93: Tag affected by secondary propagated effect according to process connectivity

The result and conclusion from the topology connectivity tool is in agreement with visual inspection of the process P&ID.

Thus, it can be concluded that the data-driven analysis has an error. A possible reason for the erroneous result is that the data analysis ignored PDA recommendation that some of the tags contained compressed data that were not suitable for analysis.

Root-cause analysis of filtered data

The clustering page shown in Figure 94 is obtained using the filtered data. The tools suggest that the time trends shown in red are similar to each other.

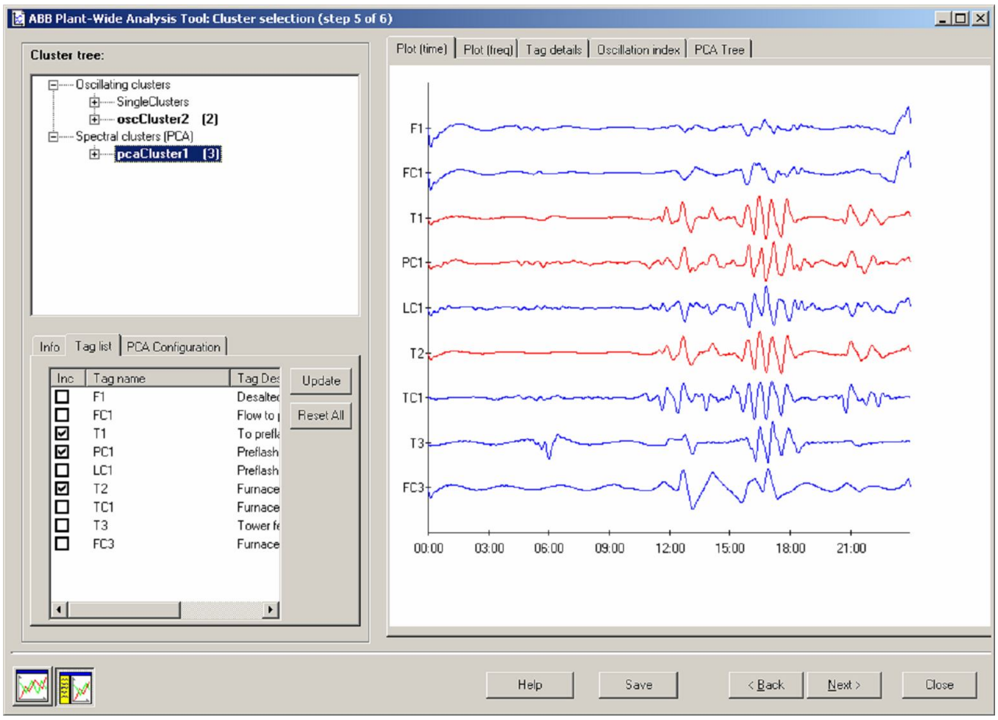


Figure 94: Clustering analysis of filtered data

However, visual inspection suggests that others are also similar hence a custom cluster is created by selecting additional tags for inclusion in the analysis: The selected tags for analysis are shown in Figure 95.

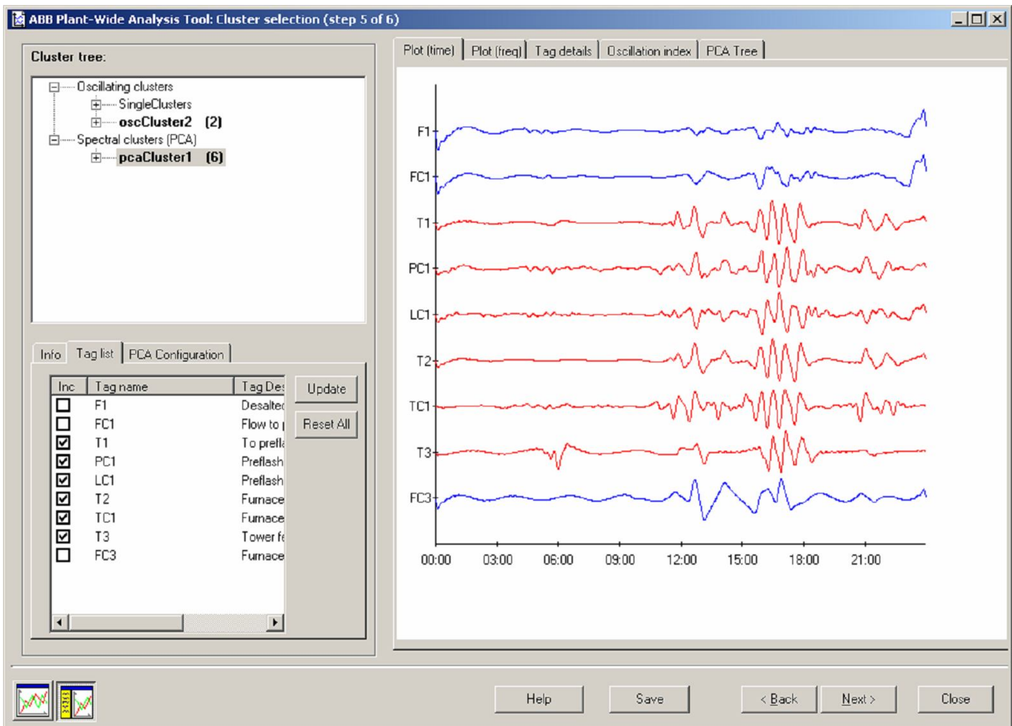


Figure 95: Custom cluster for filtered data analysis

The *Use filtered data* box is ticked when calculating the causality of filtered data shown in Figure 96.

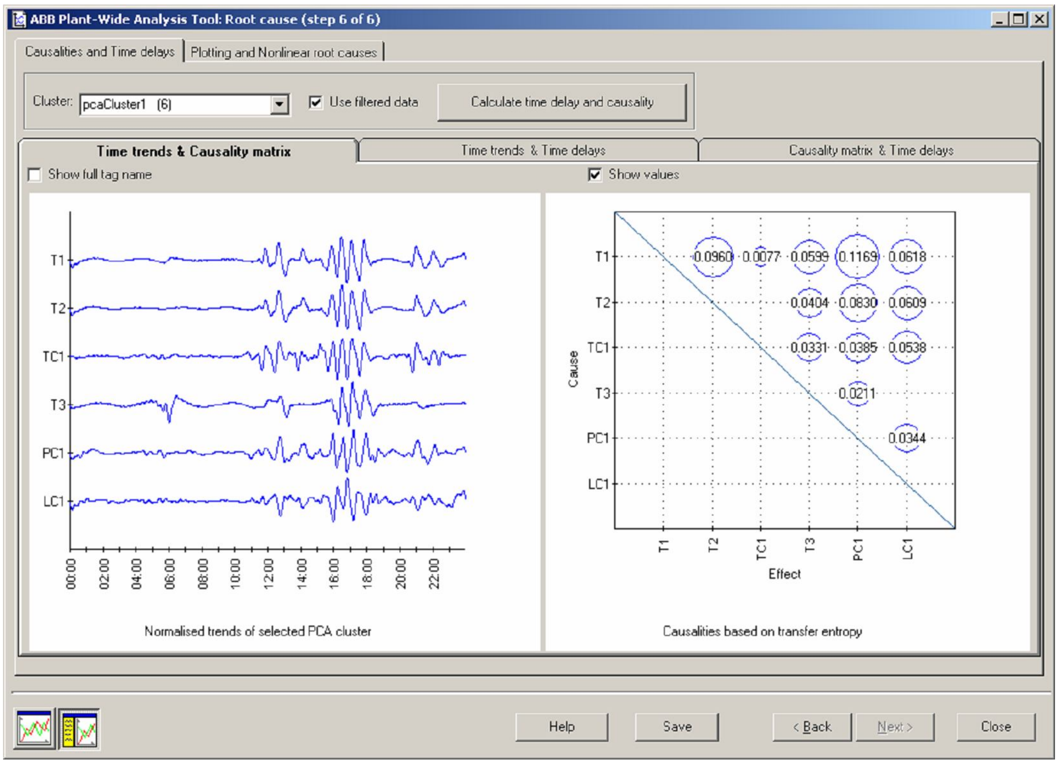


Figure 96: Filtered data-driven root-cause analysis for the atmospheric crude heating example.

Testing data-driven hypothesis against the process connectivity

The results obtained with filtered data suggests T1 is the root-cause and that the disturbance propagates to T1, T2 TC1 and T3 and then to PC1 and LC1.

The result suggested by the data-driven analysis is tested against process connectivity tool. The process plant topology connectivity analysis shows that there is feasible propagation path from the root-cause (T1) suggested by the PDA tool to other secondary propagated measurement points within the plant (T1, T2 TC1, T3, PC1 and LC1.) as shown in Figure 97.

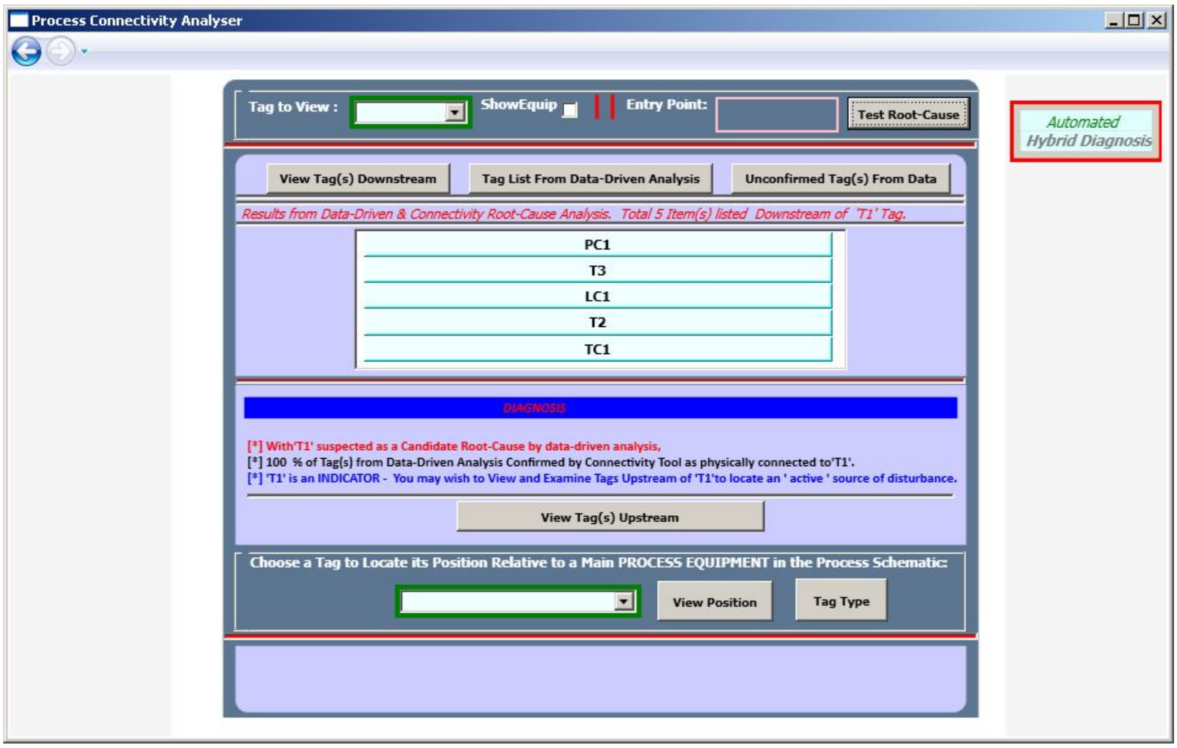


Figure 97: Data-driven hypothesis testing against process topology and connectivity

The connectivity tool indicates that 100% of the tags suggested by the data-driven analysis as measurement points affected by secondary propagated effects are physically connected and downstream of the root-cause, T1.

Visual inspection of the process schematic is in agreement with the result reported by the connectivity tool.
However, as T1 is a temperature indicator, a passive measuring device it can be concluded the real root-cause will be upstream of T1. The connectivity tool provides the functionality to view plant items upstream of a suspected root-cause candidate.

Discussion of Results

The process topology connectivity analyser would alert the data analyst to a problem with the analysis of the unfiltered data. The analyst should not have included the compressed tags when analysing the data set, however that approach would remove most of the tags of interest. The correct approach is to run the data historian with no compression for a few days and collect new data. There is no theory that says filtering overcomes the negative effects of compression, however in this case it seems to have had that effect and has yielded a more sensible result.

Further Tests and Process Connectivity Analysis

With the XML description of the process schematic loaded into the tool, the tool is ready to perform some qualitative cause-and-effect analysis of the process. Results from data-driven analysis could also be loaded for hybrid analysis. The user might want to gain insight into the process by posing questions such as:

1. Could any anomaly observed in the controlled temperature, T3 (effect) of the heated crude feeding the atmospheric tower be attributable to slave flow controller FC3 (cause)? As shown in Figure 98, the answer is yes through one propagation path. All plant items, including process equipment, are listed upon user's request.

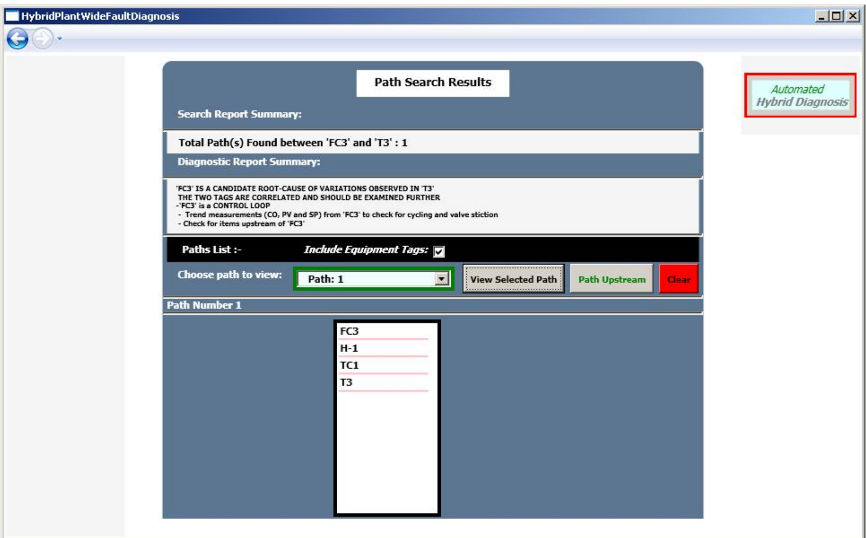


Figure 98: Path between FC3 and T3

2. By turning the query around to check if crude outlet temperature, T3 could be a cause of disturbance in slave loop FC3, the result shown in Figure 99 indicates that T3 cannot be the cause of variations in FC3.

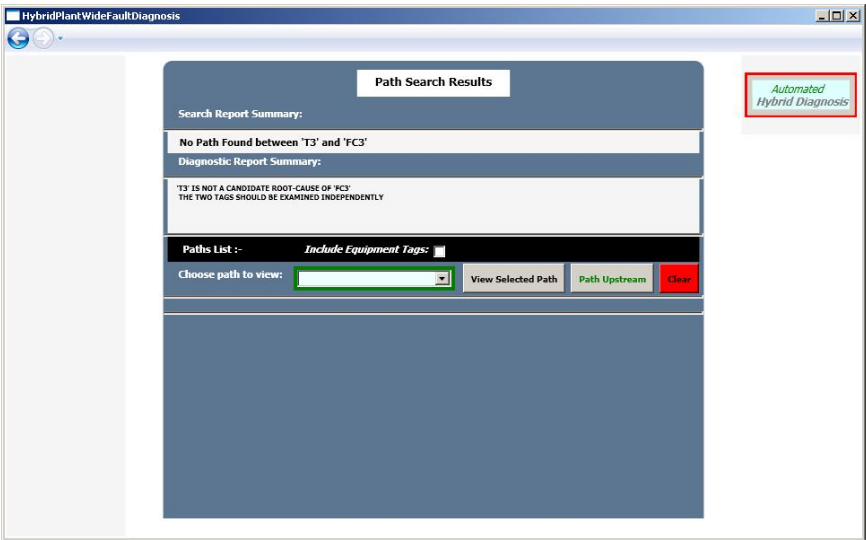


Figure 99: Path between T3 and FC3

3. The complexity increased if we were interested in finding all the plant items that could potentially affect the controlled variable T3. These items will be all items upstream of T3 and physically connected to T3 as depicted in Figure 100.

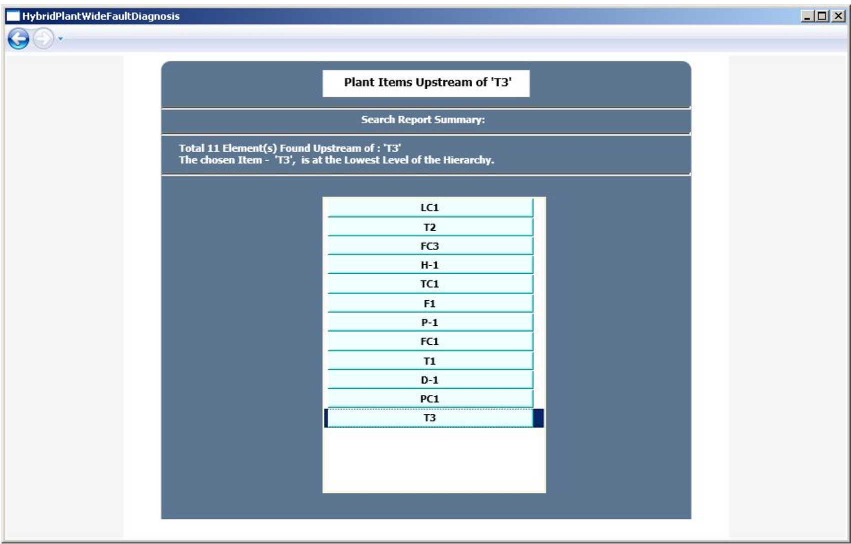


Figure 100: Plant items upstream of T3. Any malfunction in any of these elements would affect crude outlet temperature T3

4. For hybrid(data-driven and connectivity) diagnosis and other operations, such as finding the description of a tagged item and its relative position on the process schematic, the interface in Figure 101 will be used.

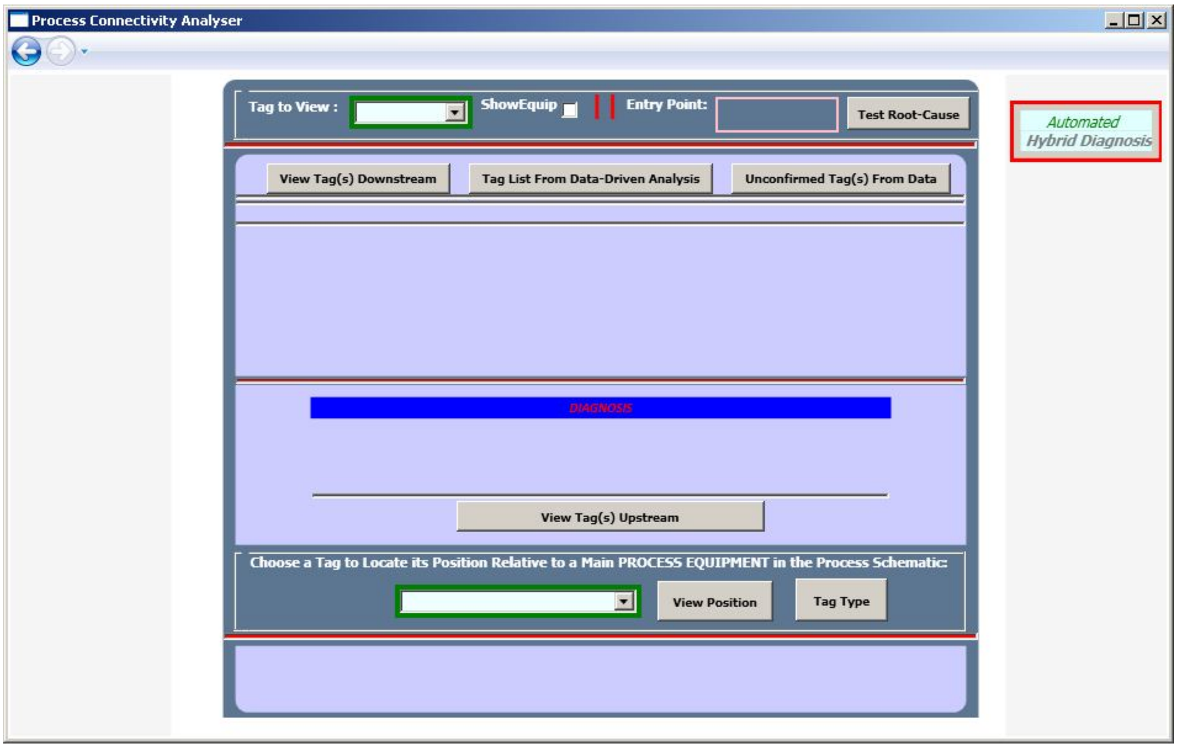


Figure 101: User interface for carrying out hybrid (combination of qualitative relationship and results from data-driven analysis) diagnosis and other operations such finding the location of an item on the P&ID and determining plant item's type.

5.5 Industrial Case Study

This section describes the industrial case study used to validate the connectivity tool. Due to the proprietary nature of the process and process plant involved, trade secrets and proprietary information have been protected in the discussions.

The case study illustrates:

1. the concept of plant-wide disturbance and
2. the use of process topology connectivity to provide insight into mechanism of disturbance propagation from a localized source.

PDA treatment of the measurements data from the plant is similar to that described in detail for the illustrative example in Section 5.4.

The case study compares results from data-driven analysis with those obtained from the process plant topology connectivity in order to validate data-driven results.

5.5.1 Process Description

The case study concerns the atmospheric crude distillation unit of Figure 102. Atmospheric distillation is one of the processes used for the separation of crude oil into straight- run cuts (fractions) by distillation under atmospheric pressure. The other process employed for crude separation into various fractions is the vacuum distillation.

The main fractions or ‘cuts’ obtained have specific boiling point ranges and can be classified in order of decreasing volatility into gases, light distillates, middle distillates, gas oils and residuum.

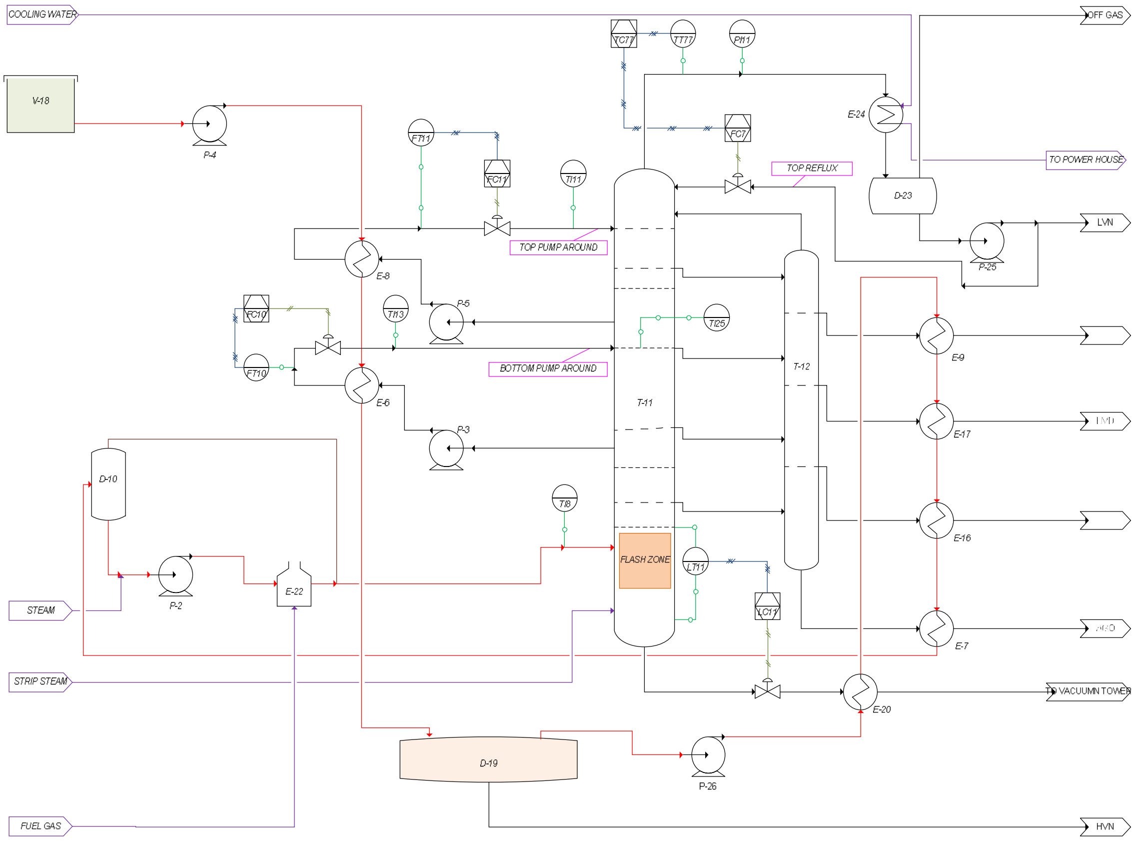


Figure 102: Process schematic for the industrial case study. The path followed by crude into the atmospheric distillation column is coloured red in the drawing. Some instrumentation have been omitted to aid visualization

Desalted crude feedstock from V-18 in Figure 102 is pumped through a series of heat exchangers E-6, E-7, E-8, E-9, E-16, E-17 and E-20 collectively referred to as a *preheat train* to raise the temperature of the crude using recovered process heat. The path followed by the crude is highlighted in red while process utilities such as steam and fuel gas lines are coloured purple on the process schematic.

The crude then flows through a pre-flash drum D-99 to separate vaporized lighter components before passing through a direct-fired feed charge pre-heat furnace E-22 where it is fed into the flash zone of the vertical distillation column T-11 at pressures slightly above atmospheric pressure and at temperatures ranging from 650°- 700° F. Heating crude oil above these temperatures may cause undesirable thermal cracking. All but the heaviest fractions flash into vapour. As the hot vapour rises in the tower, its temperature is reduced and condensation takes place. At successively higher points on T-11, the various components are separated and drawn off through side stripper T-12 as final or as intermediate products for further processing. Heavy fuel oil or asphalt residues are taken from the bottom.

Separation and collection of crude fractions take place predominantly on the horizontal trays within T-11 as illustrated in Figure 103. At each tray, vapours from below enters perforations and bubble caps which permit intimate contact as the vapours to bubble through the liquid on the tray, causing some condensation at the temperature of that tray.

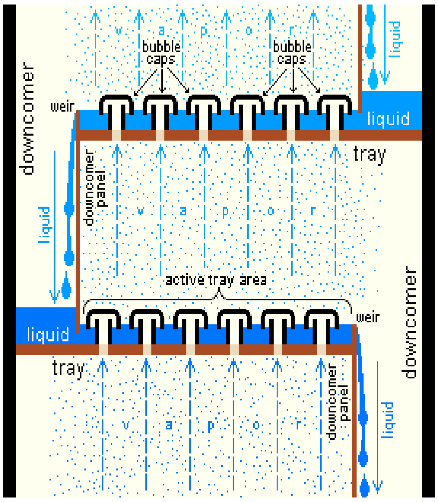


Figure 103: An example of the section of a distillation column showing details of vapour-liquid contact on bubble cap horizontal trays with liquid cross flow (source: http://www.answers.com/topic/fractionating-column)

The vapour flow pressure support the weight of the liquid on the perforated tray, preventing liquid on trays from dripping downwards (weeping). Condensed liquids from each tray drain back to the tray below through a downcomer where the higher temperature causes re-evaporation.

The evaporation, condensation and scrubbing operation is repeated many times until desired degree of product purity is achieved. Product purity can be controlled by adjusting the distillate rate or reflux rate. There are two ways to remove heat from T-11to maintain desired temperature and pressure gradients within the column and thus control product quality by cooling and condensation of upward flowing vapours. These include the use of:

- Top reflux
- Pump around-circulating reflux stream

    - Top pump around
    - Bottom pump around

Reflux pump around is operated in addition to top reflux to compensate for the lack of actual number of trays, and thus the height of the column, practically required for separation due to physical, design and economic consideration. The pump around and top reflux are operated to optimize product purification levels at the expense of reduced column height.

5.5.2 Process P&ID, XML and Connectivity

Intelligent P&ID of the process is created and exported as XML. The XML is parsed by the *Process Connectivity Analyser*. The parsed elements of the process are shown in Figure 104 interface. The figure shows the list of all the plant items, a summary of the tags as well as categorized listing into control loops, indicators and equipment.

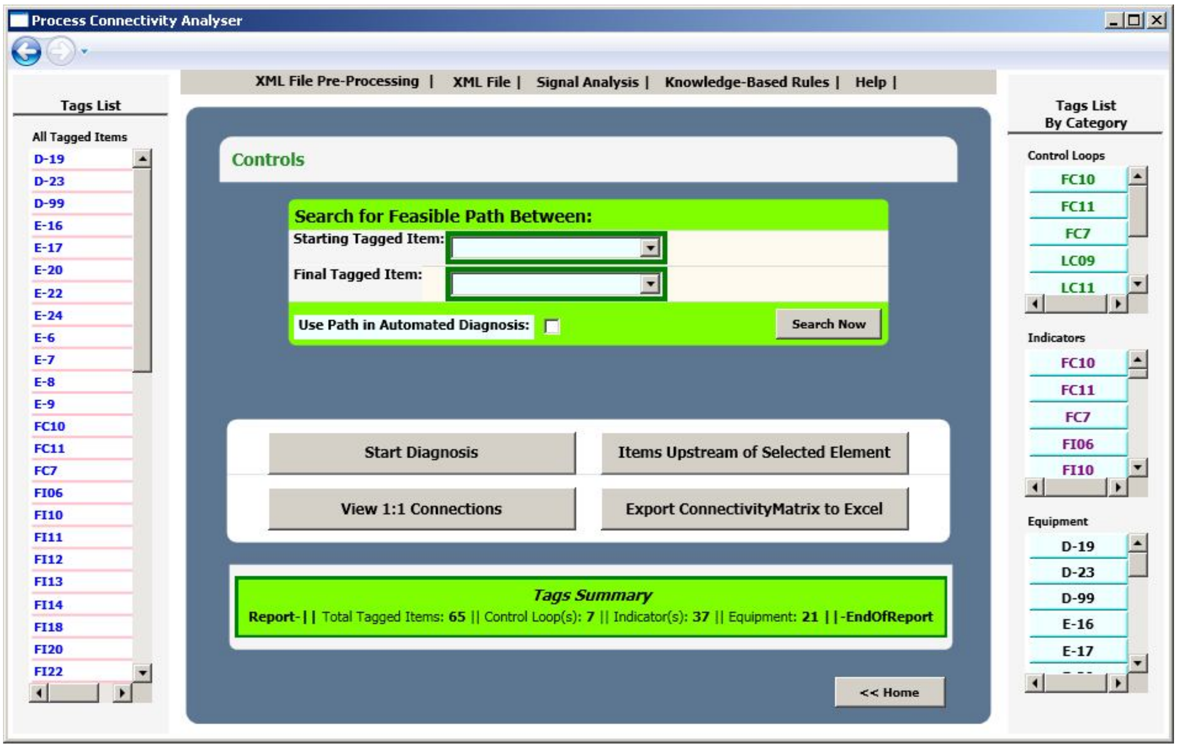


Figure 104: Plant elements of the atmospheric distillation unit as parsed from process XML description by the *Process Connectivity Analyser*

Detailed process measurement tags listing and description are presented in Table 14. The location of these measurement points are shown in the process schematic described in Figure 105.

Table 14: Process measurement tags list

Tag ID	Tag Name	Unit of Measurement	Description
1	FI12	BPH	Reflux product flow indicator
2	FI14	BPH	Side stream 1 flow indicator
3	FI8	BPH	Side stream 2 flow indicator
4	FI11	BPH	Side stream 3 flow indicator
5	FI9	BPH	Side stream 4 flow indicator
6	FI7	BPH	Top reflux flow indicator
7	FC11	BPH	Upper pump around flow controller
8	FC7	BPH	Top reflux flow controller
9	FC10	BPH	Lower pump around flow controller
10	FI22	BPH	Desalted crude flow indicator

11	FI10	BPH	Furnace charge flow indicator
12	FI20	BPH	Reduced / topped crude flow to vacuum tower
13	PC15	PSIG	Pre-flash drum pressure controller
14	PI11	PSIG	Column top pressure indicator
15	PI14	PSIG	Flash zone pressure indicator
16	TI20	DEGF	Pre-flash drum temperature indicator
17	TI12	DEGF	Top reflux temperature indicator
18	TC77	DEGF	Column top temperature controller
19	TI61	DEGF	Side stream draw 1 temperature indicator
20	TI65	DEGF	Column top tray temperature indicator
21	TI29	DEGF	TPA inlet tray temperature indicator
22	TI62	DEGF	Side stream draw 2 temperature indicator
23	TI25	DEGF	BPA inlet tray temperature indicator
24	TI14	DEGF	BPA draw temperature indicator
25	TI63	DEGF	Side stream draw 3 temperature indicator
26	TI64	DEGF	Side stream draw 4 temperature indicator
27	TI15	DEGF	Column flash zone temperature indicator
28	TI11	DEGF	TPA return temperature indicator
29	TI13	DEGF	BPA return temperature indicator
30	LC07	%	Pre-flash drum level controller
31	LC11	%	Column bottoms level controller
32	FI06	MLBH	Furnace steam flow indicator

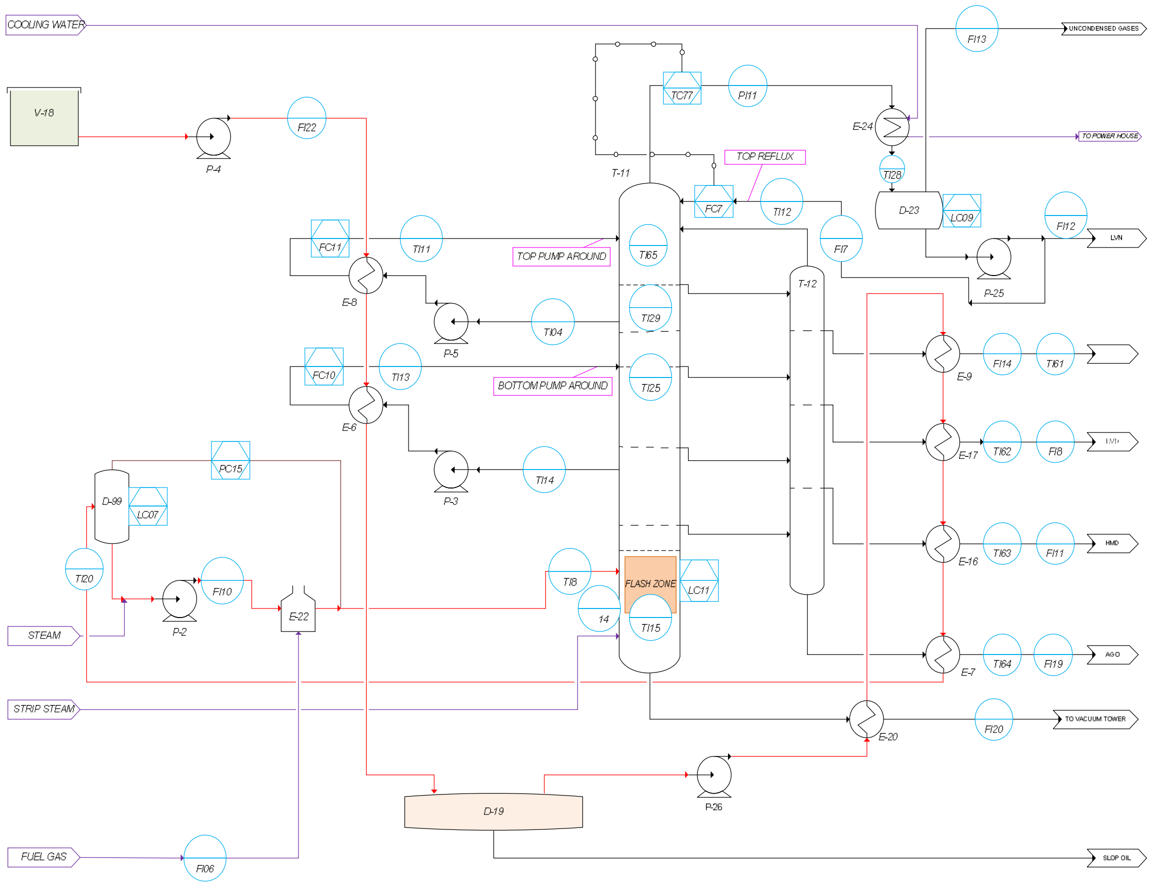


Figure 105: Process schematic showing process measurement points and control loops.

5.5.3 Process Signal Analysis

This section describes data analysis and root-cause diagnosis results obtained from signal processing using PDA tool. Detailed description of steps involved data-driven analysis of process measurements have been presented earlier in section 5.4.3.

<u>Data for Analysis and Time Trends</u>

Two days worth of minute data is analysed from archived process data historian .The signal analysis is carried out with PDA tool on straight run basis. Straight run analysis implies that no customized configuration was applied. The PDA tool was used with its default setting in the analysis. Figure 106 shows absolute values time trends of some selected tags for visualization. The plots show the tags within the column, the pumparounds and preheat sections.

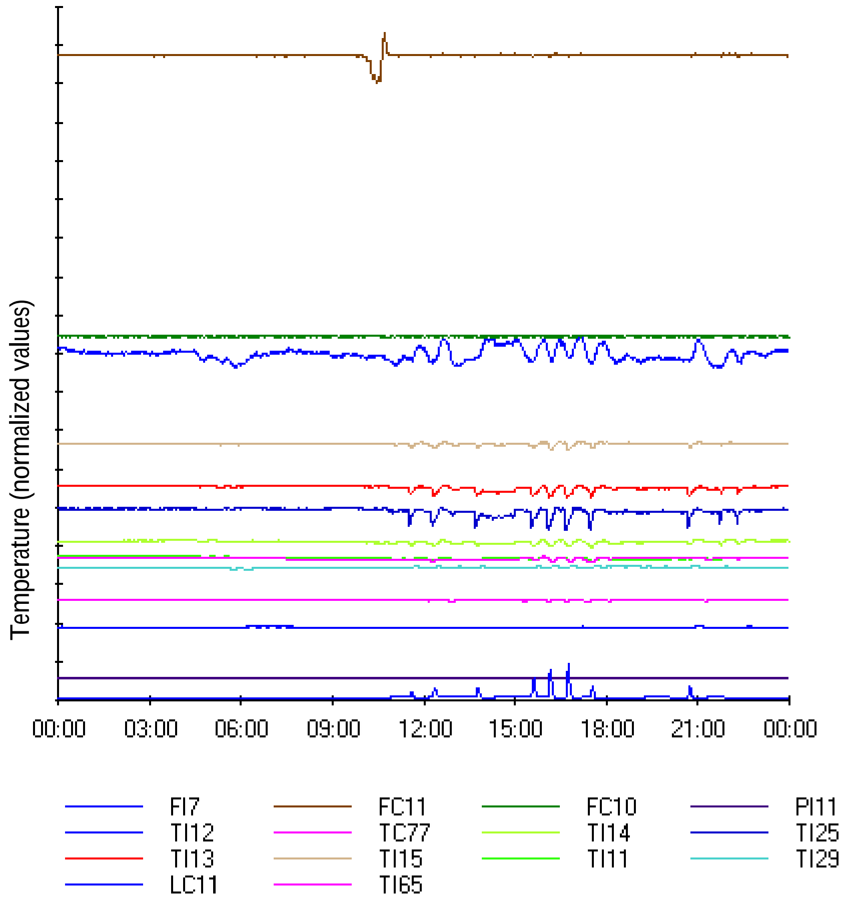


Figure 106: Time trend of absolute values of some selected tags

Visual examination of the absolute values plots indicate that deviations are most sharp and also of largest magnitude in TI25. The deviations in TI13 are relatively small, a bit later and also smoother, suggesting they are effects, not the cause. The very sharp leading edge and the large amplitude in TI25 should place it at the root-cause of the distributed upsets.

Figure 107 shows time trends and corresponding spectral components on the right hand side of process tags without filtering while Figure 108 depicts similar data visualization with filtering applied to remove noise interference, low frequencies and high frequency components from the signal spectra.

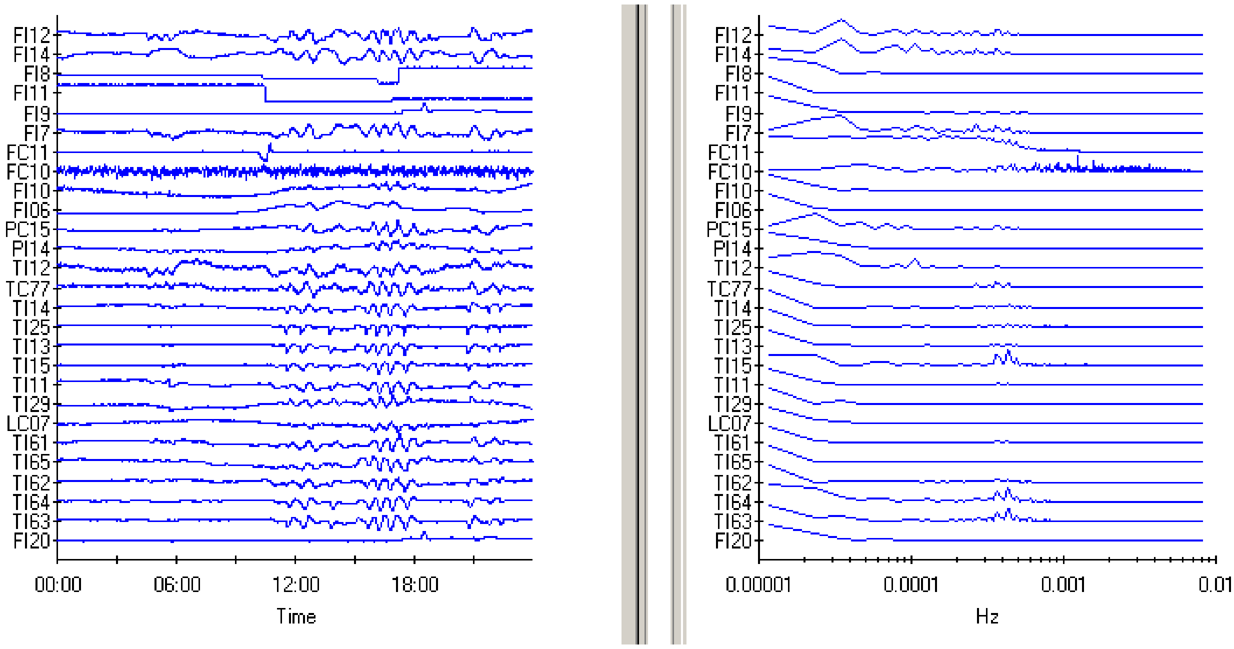


Figure 107: Time trend and spectral of uncompressed and unfiltered tags

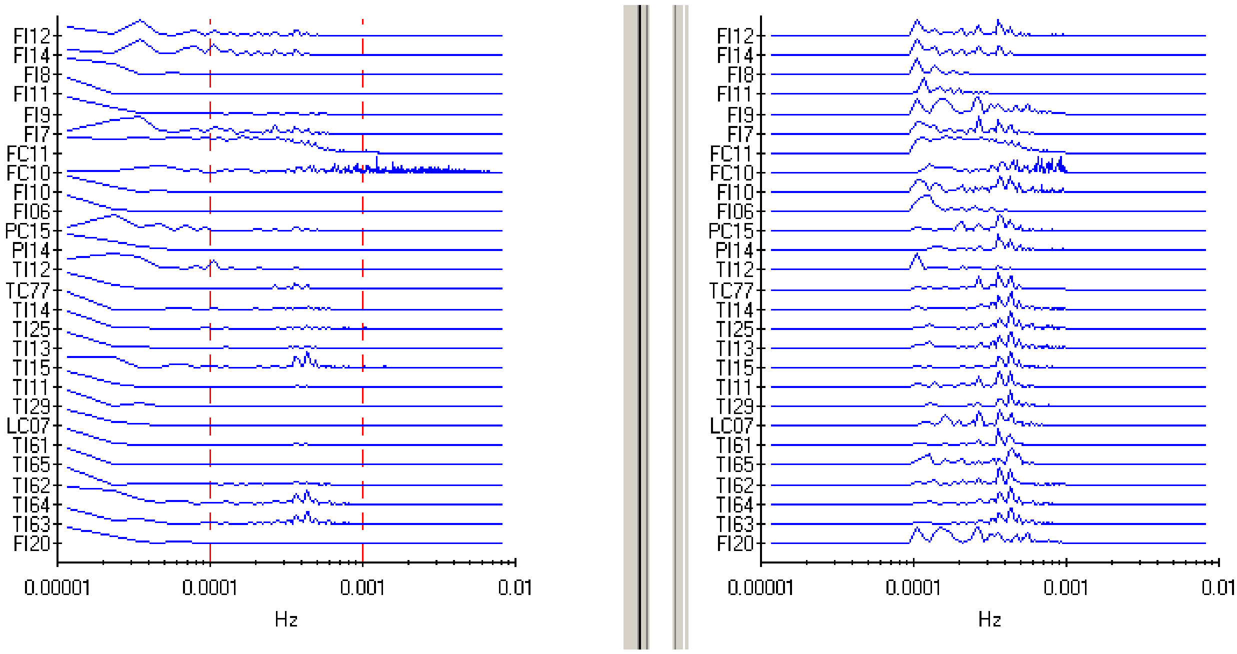


Figure 108: Spectra with band pass filter 0.0001Hz and 0.001Hz

With the application of a band-pass filter, the spectra in Figure 108 (right hand panel) reveals co-ordinated spectral peaks at around 0.0004Hz suggesting plant-wide oscillation among some process variables. A view of the time domain visualization of filtered signal in Figure 109 reveals that some tags are oscillating in tandem, just as observed in the frequency domain.

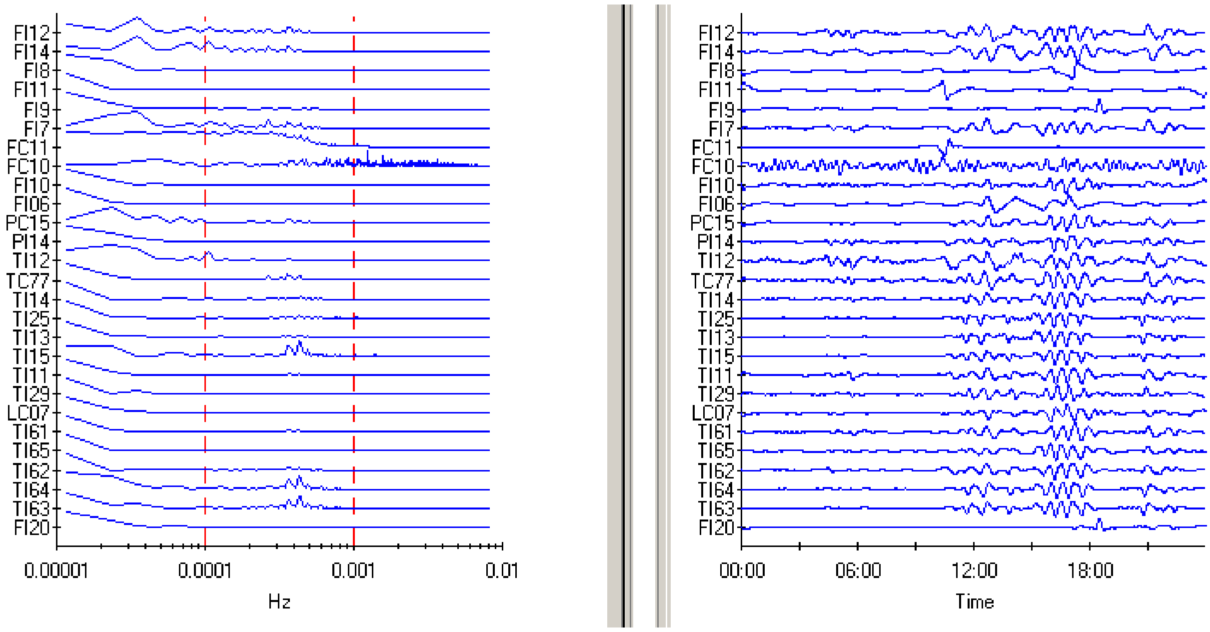


Figure 109: Time series with band pass

The PDA tool is able to characterise and present the tags found to be participating in plant wide oscillation with coordinated behaviour as a cluster. The clustered tags are shown in Figure 110 as red lines.

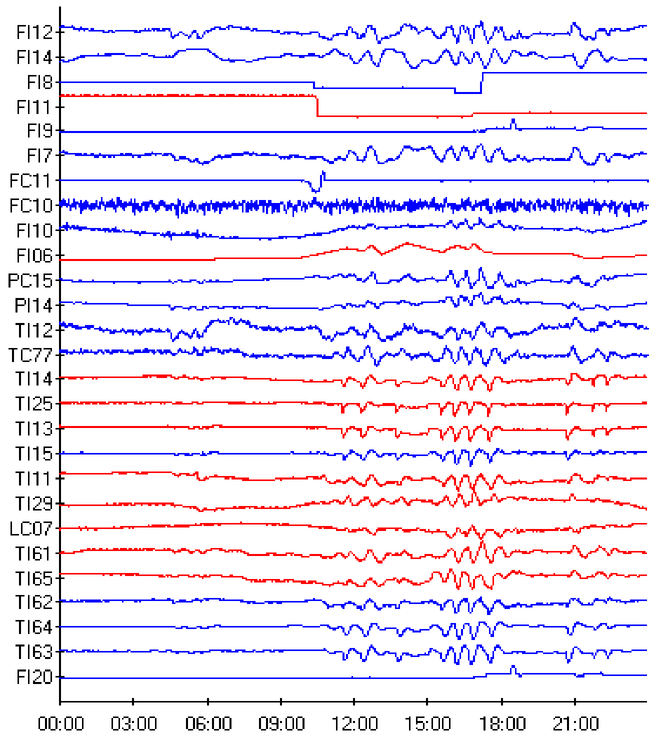


Figure 110: Straight run PDA analysis. PDA tool clustered ten tags-red lines as participating in plant-wide disturbance based on spectra similarity

Two main methods are used to highlight the presence of possible periodic and non-periodic disturbances.

- A spectral cluster analysis is based on the automated comparison of the spectra for detecting similarities, hence it groups tags with similar spectral features. The method used is the spectral principal component analysis as detailed in (Thornhill, *et al*., 2002). Comparing spectra for detecting similarities may be done visually in small scale cases with a small number of tags. In larger scale cases the automated spectral clustering method becomes a necessity.
- A second method looks for clusters of oscillating measurements. The output is a list of clusters of tags characterized by their oscillation period. The oscillation detection uses signal processing methods described in (Thornhill, *et al*., 2003).

It will be noted that this result is obtained by operating the PDA tool in its default setting as visual examination shows that some plots signature appear out of place while some seemingly similar tags have not been grouped with the cluster. However, as stated earlier on, the default settings of PDA will be used in the analysis. Figure 111 shows the plots of time trends of the clustered tags

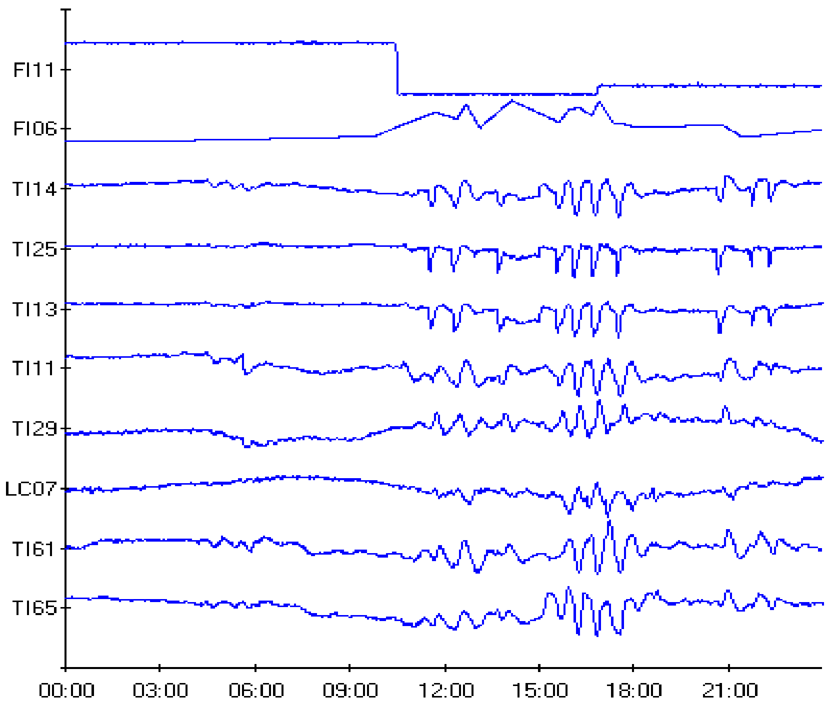


Figure 111: Time trends of the ten clustered tags

The location of the tags suspected to be participating in a plant wide oscillation is depicted in Figure 112 as red spots placed manually. The disturbance distribution indicates that the tags within column as well as tags around the column are affected.

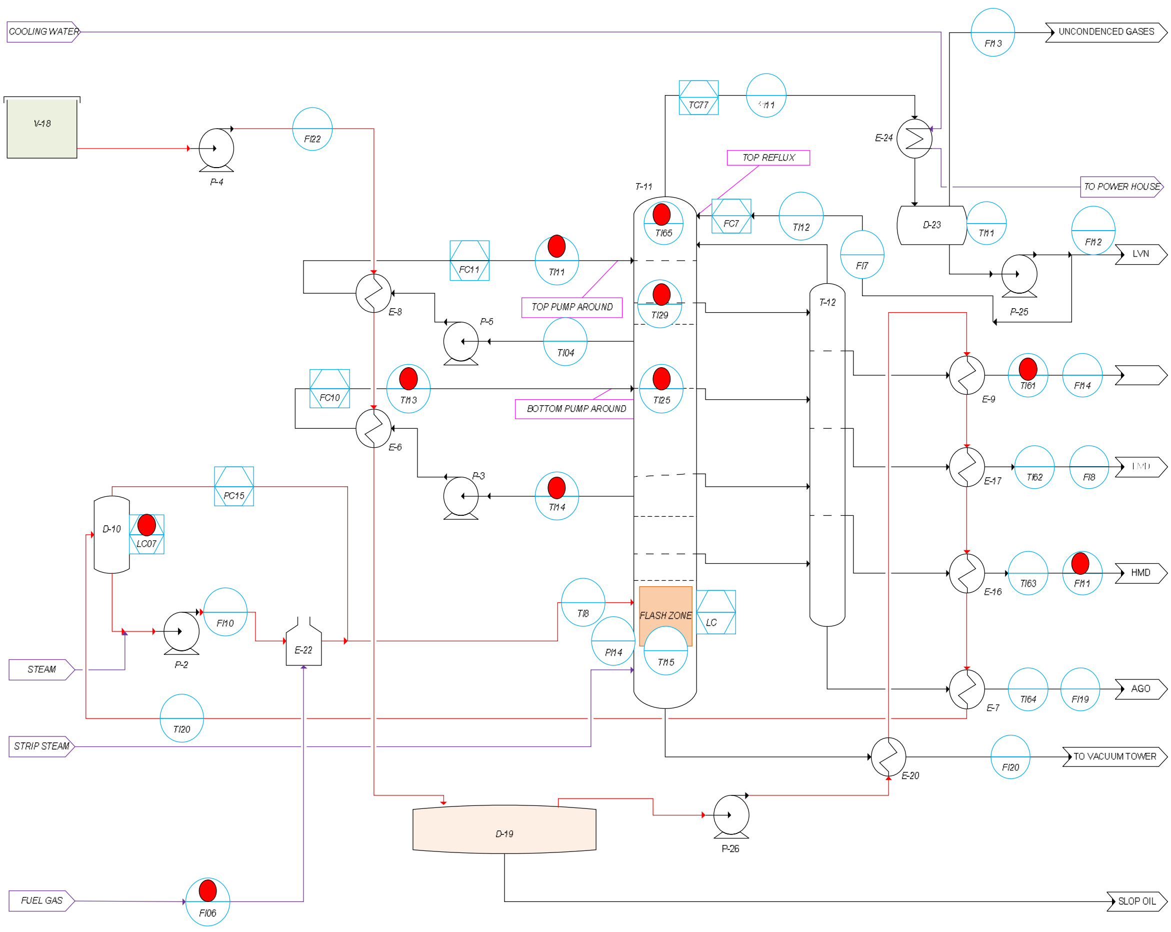


Figure 112: Location of clustered tags on the process topology

Data-driven Root-cause Diagniosis of Process Measurements

The conclusions reached by the PDA tool on the possible root-cause candidate among the clustered tags are depicted in Figure 113.

There are several tools to help with root-cause analysis in the PDA tool. General purpose tools include causal analysis based on time delays between measurements (Bauer and Thornhill, 2008) and a probabilistic assessment called Transfer Entropy (Bauer, *et al.*, 2007) which is more sensitive than time delays alone in detecting causality.

Root-cause analysis also examines non-linearity in the measurement time trends (Thornhill, 2005), where a non-linear time series means one which is generated by a non-linear system. Present values of a non-linear time series have a non-linear dependence on past values. The time trend within an oscillating cluster having the highest nonlinearity index can be interpreted as the root-cause of the oscillating disturbance. The reason is that process plants act as a low pass filter, i.e. a tag close to a source of nonlinearity will be more nonlinear than another tag further from the source. Examples include control loop limit cycles caused by sticking valves and fluid dynamic instability such as slugging.

In Figure 113, the top lefthand side panel shows the normalized time trends of the clustered tags. The plot in the right hand side of the top panel shows the causality based on transfer entropy.

Figure 114 shows an advanced display from the PDA tool in the form of a hierarchical tree constructed from an analysis of the data set using spectral PCA.

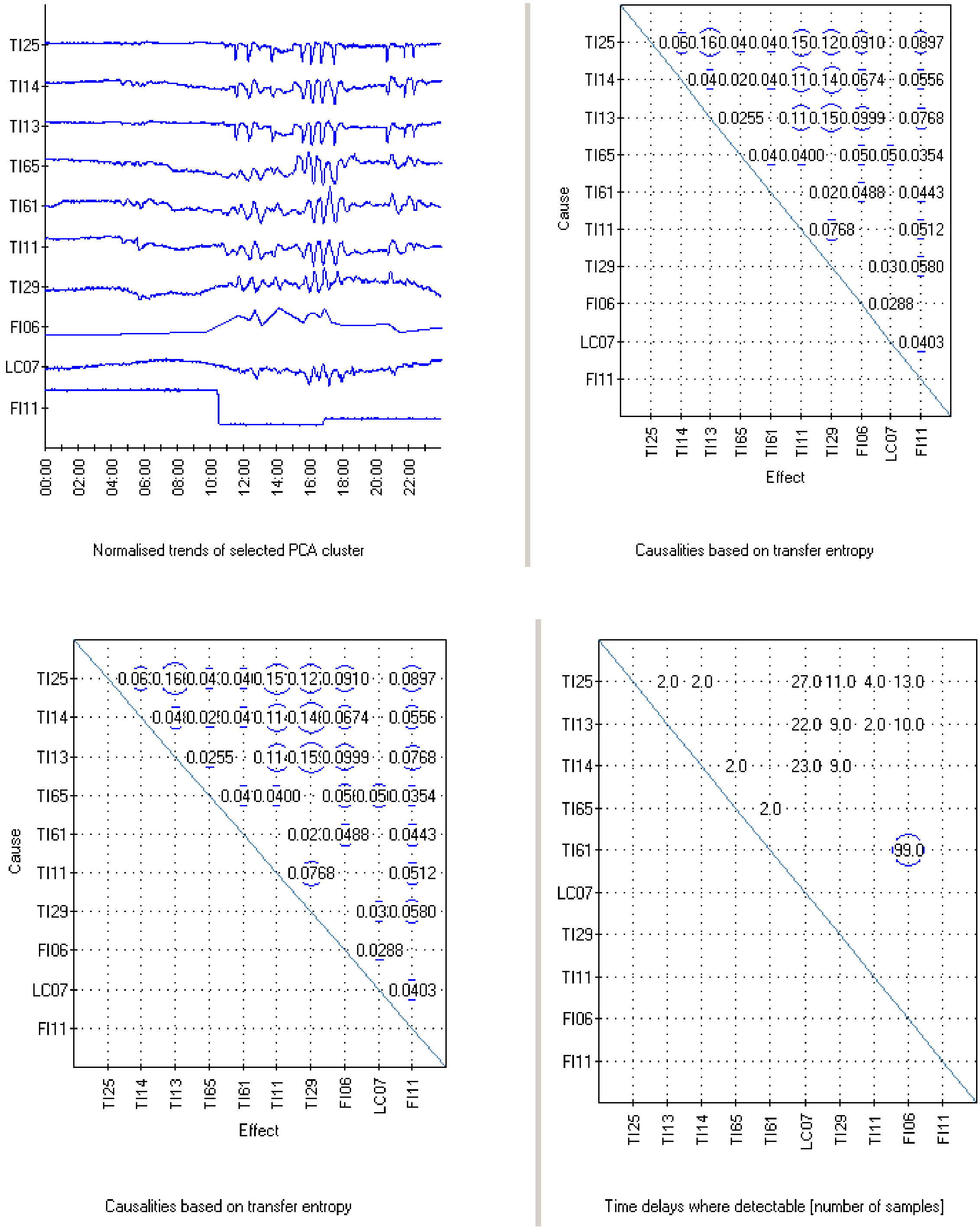


Figure 113: Time trends and causality analysis of measurements cluster. Causality analysis suggests the order of events in the process. The results displayed implies that changes in TI25 are causes of changes in the other tags and that FI11 is the last in the causal chain

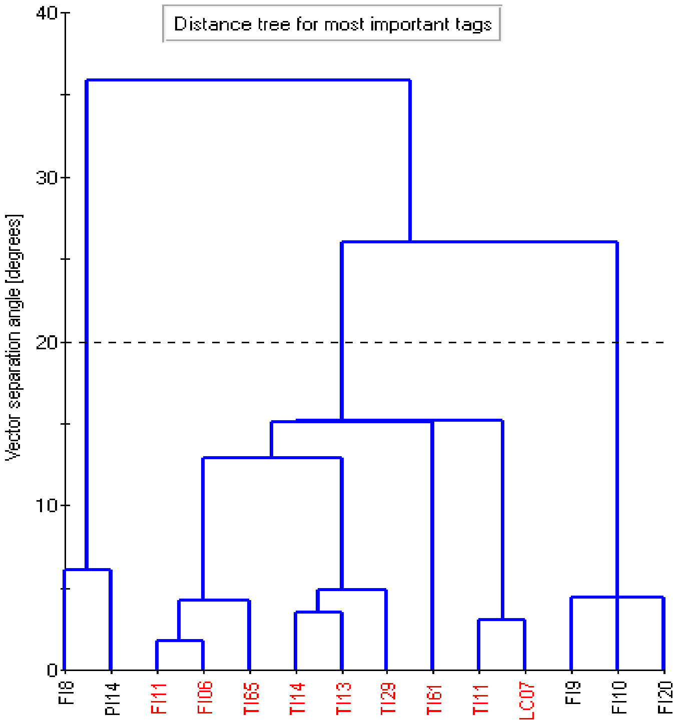


Figure 114: Hierarchical spectral classification tree for the tags in the atmospheric distillation unit

The height of the highest horizontal line that has to be followed to get from one tag to another tag is a measure of how similar their frequency spectra are. Low horizontal connection-lines indicate similar spectra. Clusters look like leaves on the end of branch of an inverted tree.

5.5.4 Process Connectivity and Directionality Information from P&ID

The process connectivity matrix is a necessary step in using process connectivity and directionality for analysis. The process connectivity matrix is generated from an XML file compliant with ISO15926 data structure. The process connectivity matrix as automatically generated by *Process Connectivity Analyser* tool is shown in Figure 115.

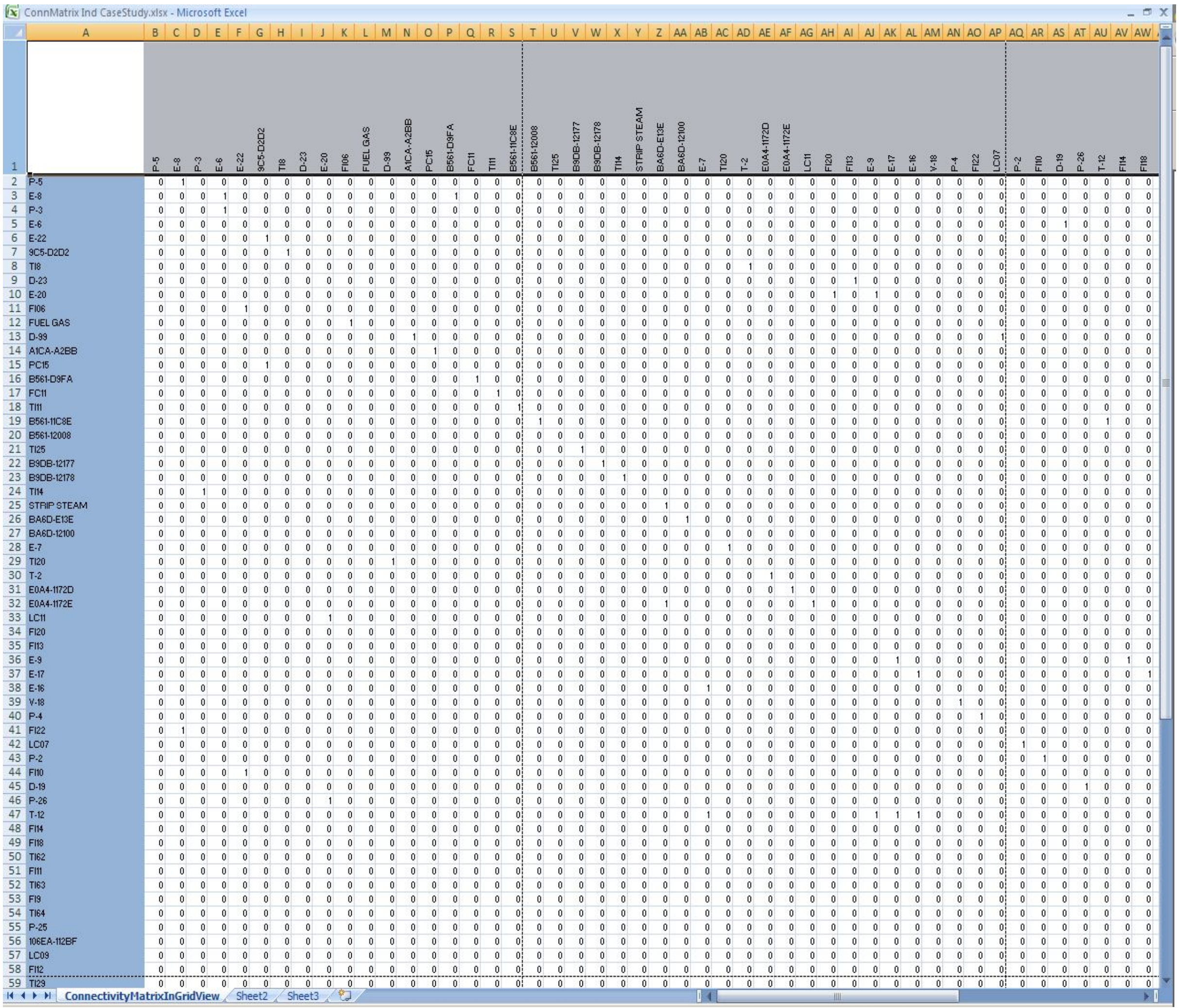


Figure 115: A portion of connectivity matrix of the atmospheric distillation case study automatically generated by *process connectivity analyser*

The complete matrix of the refinery unit case study is a 91 by 91 matrix and is much larger than previous studies (Scherf, 2006; Thambirajah, *et al*., 2009; Yim, *et al*., 2006). The functionality of the *Process Connectivity Analyser* which allows multiple P&ID drawings to be merged for analysis is a demonstration of the scalability and industrial relevance of the research concepts.

The connectivity matrix shows one to one connections and direction among all plant items depending on the entry, one for connectivity and zero otherwise, at the intersection of a row and column matrix element.

Hypothesis Testing with Process Connectivity Analyser

This section compares the results obtained from pure data-driven analysis against the process topology to ensure that the conclusion reached makes sense or otherwise.

Figure 116 shows the output from *Process Connectivity Analyser* tool which combined process connectivity and directionality information with the results from data-driven analysis.

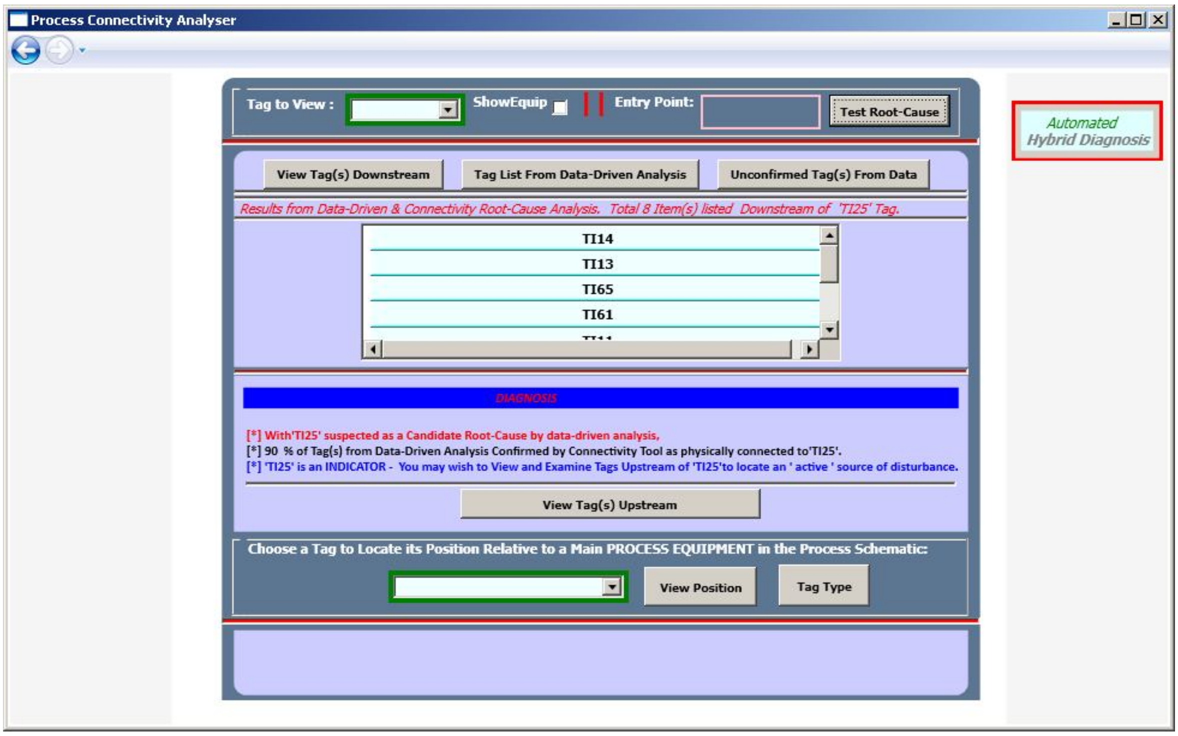


Figure 116: Data-driven analysis hypothesis testing for root-cause using process connectivity and directionality information. Connectivity tool found one spurious tag among data analysis cluster

The results shows that of the ten tags suspected to be exhibiting co-ordinated behaviour, only nine of such tags, representing 90% of the cluster are physically connected.

Only eight tags are found to be affected by TI25 due to secondary propagated effects from the suspected root-cause according to data-driven analysis as oopposed to nine suggested by the PDA tool. Figure 117 displays the spurious tag, FI06 included in the data-driven analysis PDA data analysis tool. The tag is encircled in green for easy identification on the process topology in Figure 118.

Figure 117: Spurious result from data-driven analysis detected by the *Process Connectivity Analyser* tool. FI06 is not directionally connected to the suspected root-cause

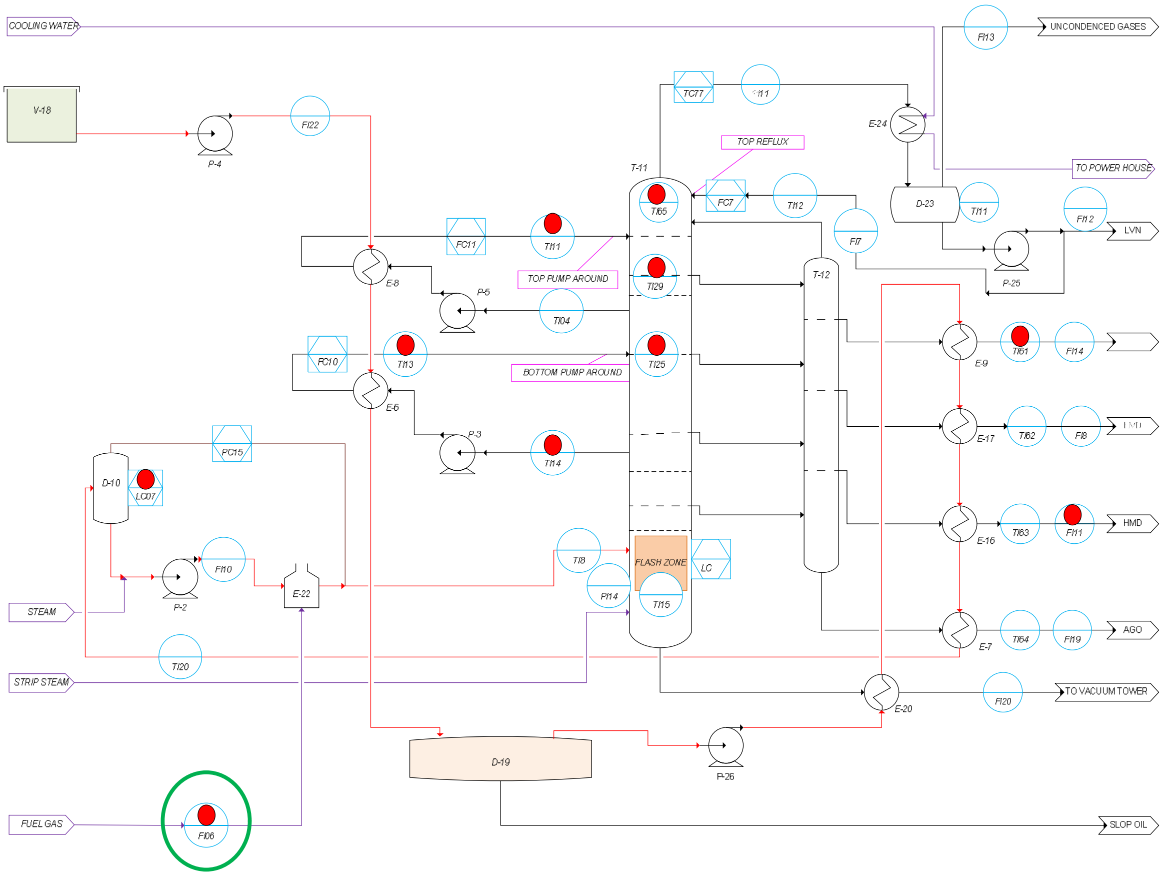


Figure 118: Location of suspected spurious result by signal analysis on the process schematic in green circle

Visual inspection of the process schematic shows that FI06 is the flow measurement indicator of the fuel gas utility to the feed preheat furnace.The direction of flow of the fuel gas indicates that it can only be a cause and not effect to disturbance in tag TI25 originating from within T-11 atmospheric fractionating tower.

The *Process Connectivity Analyser* tool allows further querrying and informationm gathering useful for diagnostic purpose on the suspected root-cause. An example is shown in Figure 119 which provides additional information about FI25. For example, the fact that TI25 is an indicator implies that it is not likely to be the disturbance generator as an indicator is only a meaurement point. One may have to look upstream for a control loop or an equipment as the likely culprit.

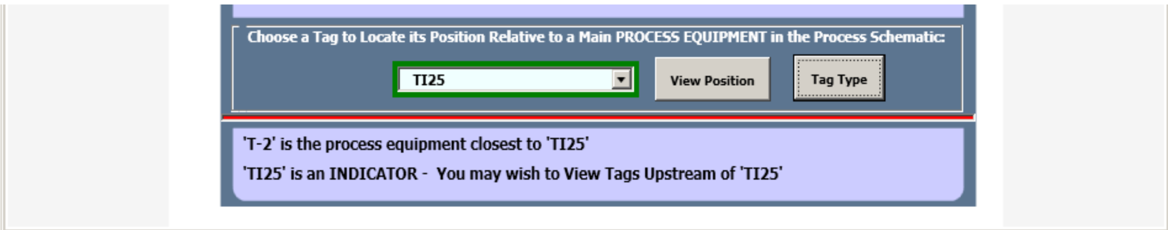


Figure 119: More information about the suspected root-cause-location and type

Root- Cause Analysis

The suspected root-cause from data-driven analysis tested against the connectivity tool is TI25 tag. Further examination of this tag indicates that it is an indicator used for measurement. However, since data-driven analysis only considers measurement points, it is incapable of identifying active disturbance generator without recourse to the physical process topology to look for plant items upstream of the suspected measurement point as the root-cause.

The red dotted circle in Figure 120 captures the first three plant elements upstream of TI25. TI13 is another temperature measurement indicator and thus a passive element. The first active plant item, capable of inducing process disturbance is the control loop FC10.

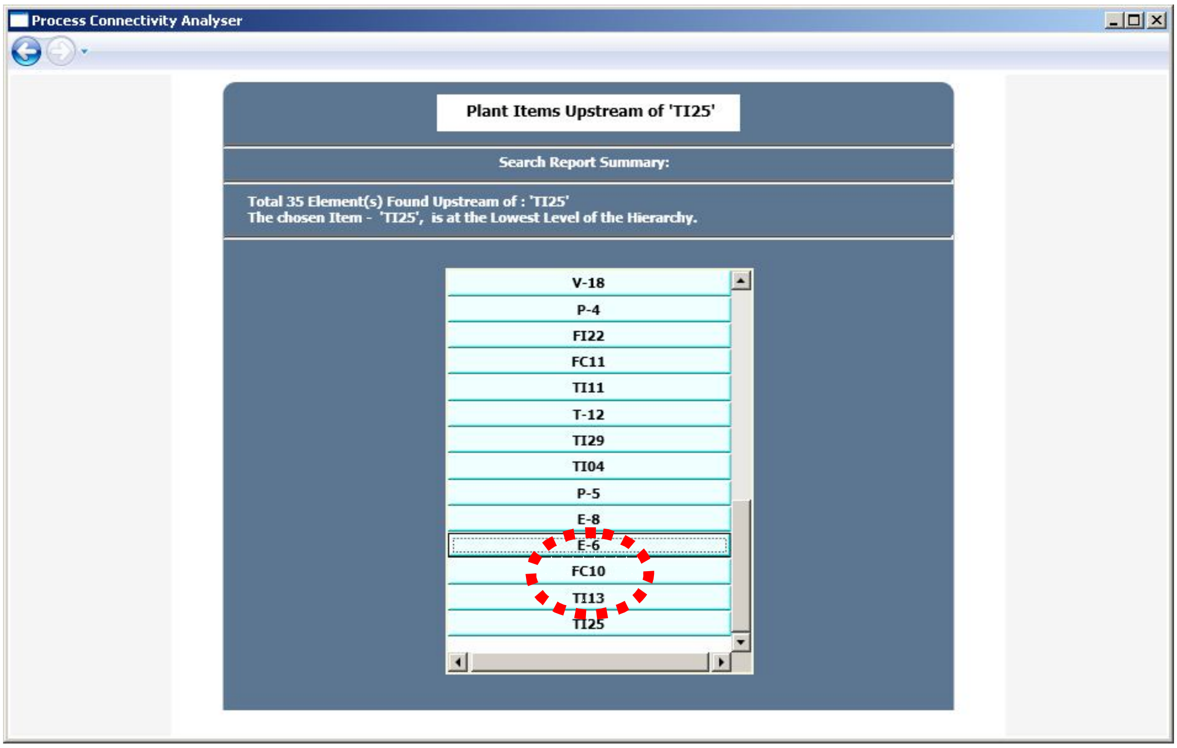


Figure 120: List of plant items upstream of the TI25.The first active plant element is control loop FC10 and heat exchanger E-6. TI23 is an indicator.

FC10 is not, however, detected by data analysis to be affected by secondary propagated effects from the root-cause; hence it was not included in the cluster. The next item on the list is the heat exchanger E-6. This is a shell and tube type heat exchanger with cold crude flowing counter current to the bottom pumparound reflux. Cooling of the hot reflux is achieved by heat exchange with cold crude before returning the reflux to the tower. This process ensures that heat is recovered and utilized for higher plant operational efficiency. Other plant items upstream of TI25 are listed in Figure 120.

Checks on Feasible Propagation Paths

The *Process Connectivity Analyser* tool can be used to validate all secondary propagation effects from the root-cause as shown in Figure 121, Figure 122, Figure 123 and Figure 124.

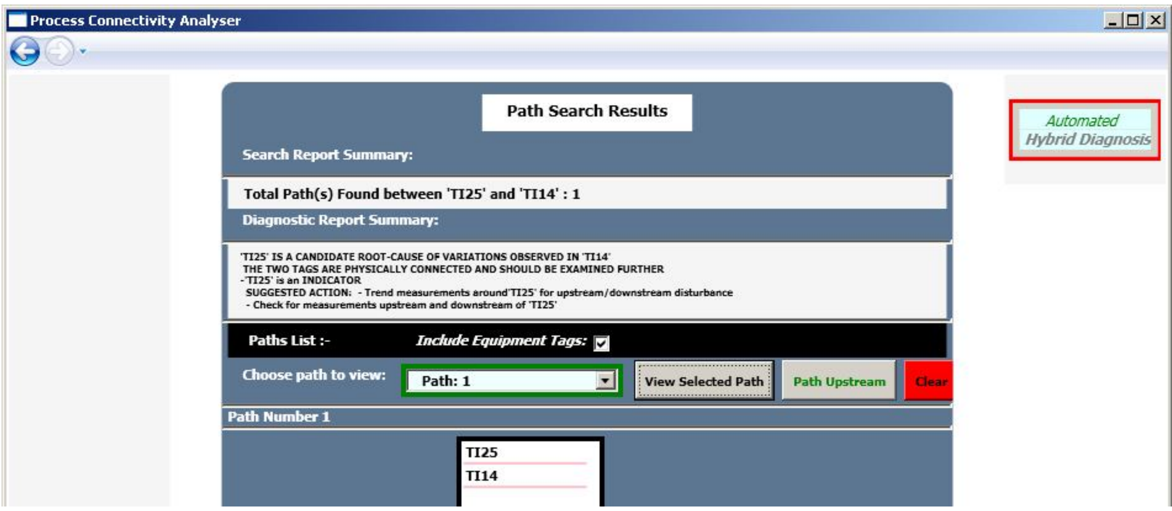


Figure 121: Propagation from *TI25* to *TI14*. There is direct connectivity between two plant items

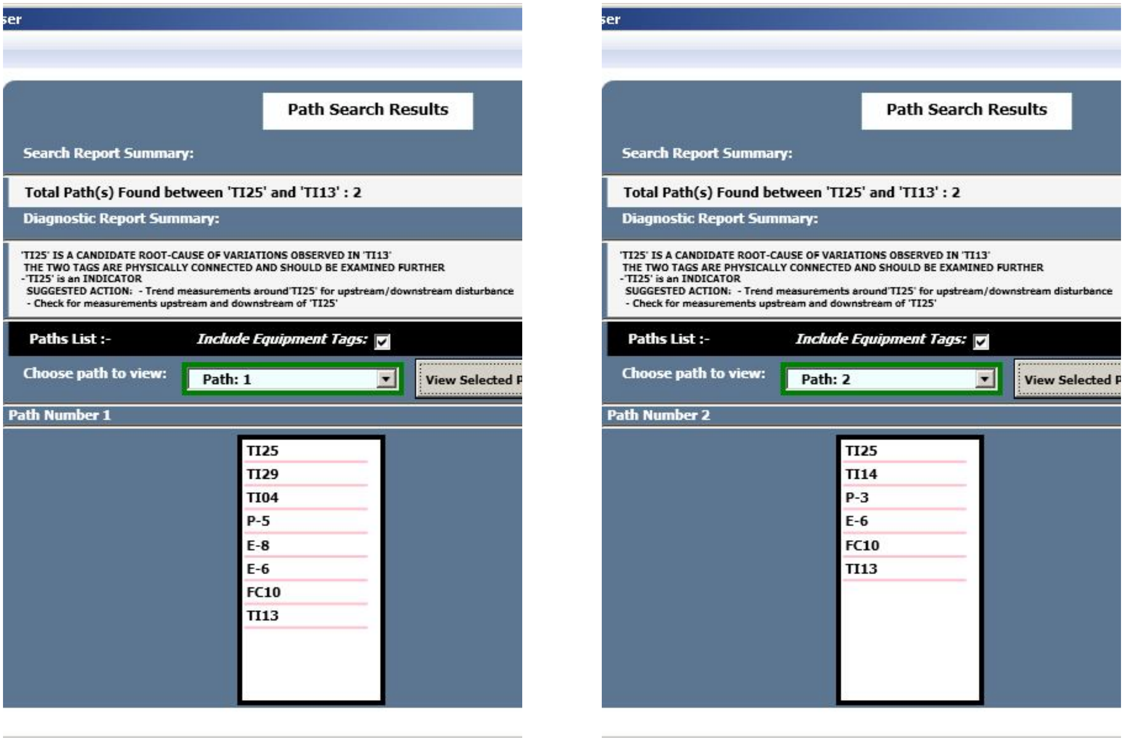


Figure 122: Propagation from *TI25* to *TI13*. Process connectivity analyzer shows that there are two feasible propagation paths listed in the two panels above

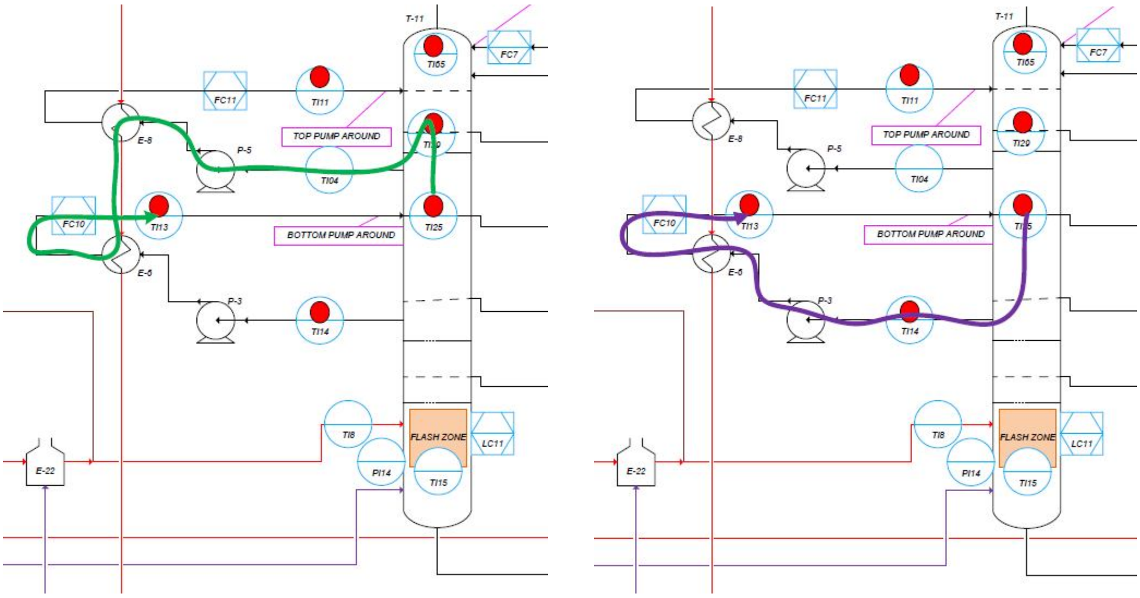


Figure 123: Disturbance propagation path tracing visualisation from *TI25* to *TI13.* Path 1 is shown on the left hand panel in green. Path 2, coloured purple, is depicted on the right hand panel

Similar analysis can be carried out of other secondary affected tags. For example an analysis carried out to find out physical connectivity from suspected root-cause element *TI25* to *FI06* depicted in Figure 124 shows that there is no feasible propagation path from *TI25* to *FI06* tag which was included in the root-cause analysis results from pure data-driven- analysis.

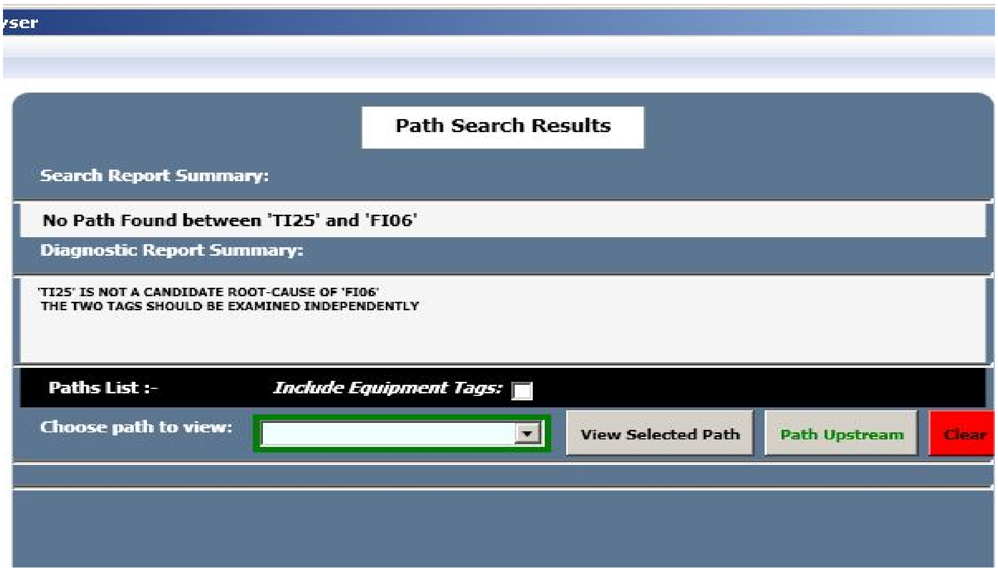


Figure 124: Propagation from *TI25* to *FI06.* Process connectivity analyser shows that there is no feasible propagation path between the two tags

5.5.5 Discussion of Results from Industrial Case Study

Incorporation of information from the process topology has provided deeper insight into the process diagnosis and helped in elimination of spurious result from the data-driven-analysis alone. One of the secondary propagation measurement points, *FI06*, included in pure data-driven analysis is found to spurious according to process topology. This was detected by the *Process Connectivity Analyser* tool with the majority, eight out of nine, of the tags in the cluster confirmed as physically connected to suspected root-cause.

Further analysis showed that the likely real cause of the disturbance is the heat exchanger *E-6* where cold crude and bottom pumparound exchange heat energy. However, since *FC10* did not participate in the coordinated plant wide oscillation, it can be inferred that temperature spike observed at *TI25* within the column is due to composition changes. This can be caused by occasional leakage between the cold crude and the pumparound reflux which introduces heavier components into the upper section of the column where the bottom pump around enters the column.

This conclusion was considered logical by process control engineers but this cannot be confirmed until the plant undergoes the next scheduled turn around maintenance to confirm the existence of a leakage within the heat exchanger.

Disturbances caused by external elements such rain shower can be ruled out because no temperature spike disturbance was observed in other columns in the neighbourhood of the unit under consideration within the plant.

5.6 Section Summary

The section has completed an open run PDA analysis of the measurement data from atmospheric distillation unit of a refinery. An open run of the PDA tool means that all analyses are run at default settings to find major disturbances with minimum amount of interaction with the user.

The PDA tool found a major cluster consisting of ten measurement tags and placed a temperature indicator within the column as the root-cause of distributed disturbances affecting nine other tags within the plant. The data-driven hypothesis was tested against the process physical layout using the *Process Connectivity Analyser* developed in this project. The *Process Connectivity Analyser* suggested that the result obtained by the PDA tool is plausible and the suspected cause of observed distributed plant-wide disturbance identified.

5.7 Academic Case Study

This section presents the electronic description of the Tennessee Eastman (TE) control challenge, a widely accepted test bed for implementing process control strategies (Lyman and Georgakis (1995); Downs and Vogel, 1993).

The ISO 15926 compliant XML representation generated from intelligent P&ID of the TE process provides the starting point for a range of automated analysis methods including the extraction and manipulation of connectivity information. As far as can be ascertained, this is the first time that this challenge process has been represented in such a machine-readable format coupled with the connectivity matrix and the manipulations associated with it such as a reachability matrix discussed earlier in Section 2.3.8 of this thesis.

To demonstrate the efficacy of the tool developed in this thesis, the TE connectivity information is used to test the hypothesis about disturbance propagation based on the TE plant-wide control structure recommended by Lyman and Georgakis (1995) and implemented by Chiang and Braatz (2003). The application will also demonstrate the versatility of the connectivity tool developed in this work and show that the tool is not tied to a particular data-driven analysis tool such as the PDA used extensively in this work.

5.7.1 Process Description

It has always been a desirable goal within process control community to have a realistic, standard, robust and industrially relevant process control test bed to evaluate the various process control strategies proposed in the academic environment and several papers have been published on the subject (Prett and Morrari, 1986). The classic control challenge of the Tennessee Eastman process (Downs and Vogel, 1993) shown in Figure 125 is used as the case study in this academic illustration.

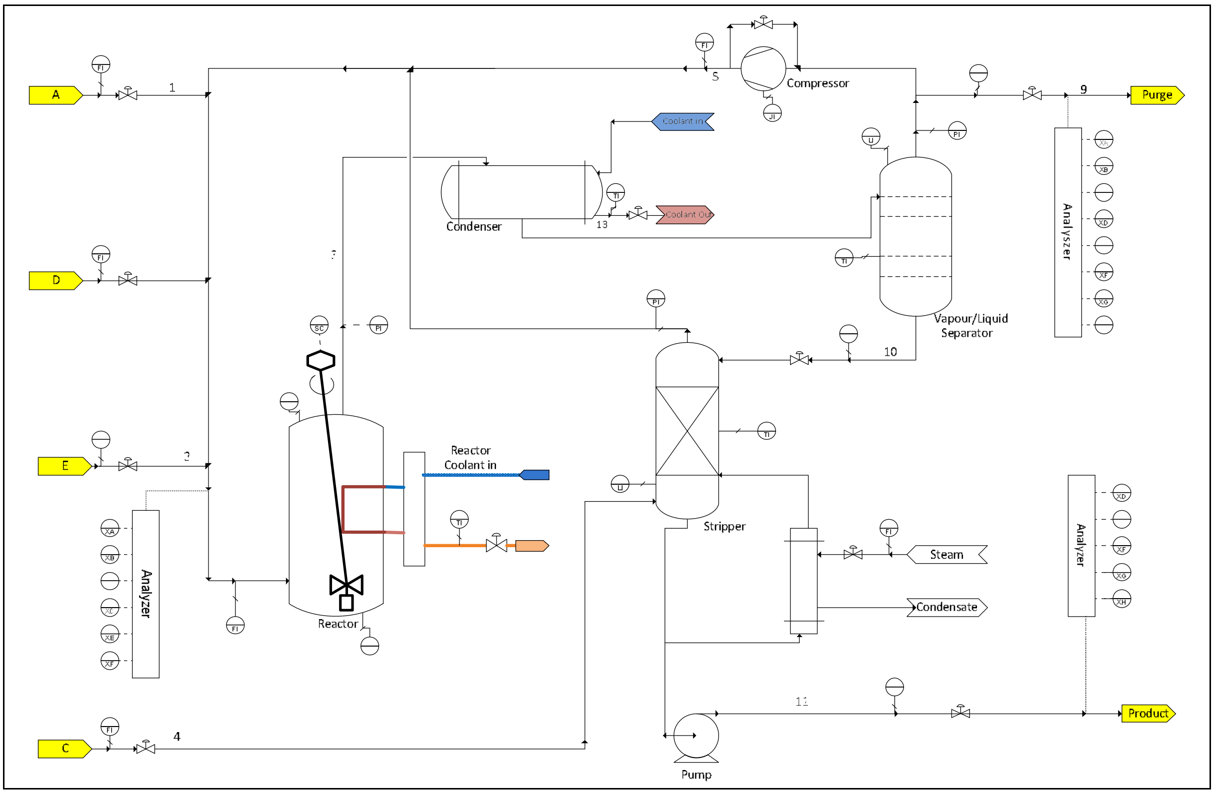


Figure 125: The TE process schematic

The process consists of five main processing units: a reactor, a product condenser, a vapour-liquid separator, a recycle compressor and a product stripper.

Two liquid products G and H are produced from four gaseous reactants A, C, D and E. A recycled inert gas is added to the reactants stream and fed to the chemical reactor unit where exothermic and irreversible reactions of the reactants take place. The order of the chemical reaction with respect to the concentrations of the reactants is approximately first-order. There are also an inert and a by-product making a total of eight components: A, B, C, D, E, F, G, and H. Detailed description of the process and chemical kinetics are discussed by Downs and Vogel (1993) and Lyman and Georgakis (1995). The tags description used to produce the TE CAD drawing is shown in Table 15.

Table 15: Tags list used in the TE CAD schematic and intelligent P&ID

Equipment		
Tag	**Description**	**Remarks**
• CD-01 • CP-01 • PM-01 • PM-02 • RX-01 • SP-01 • SR-01 • XE-01 • IM-01	• Condenser • Compressor • Pump (product) • Pump (stripper supply) • Reactor • Vapour/liquid separator • Stripper • Stripper recycle • Reactor impeller	
Control loop		
Tag	**Description**	**Remarks**
• FC1 • FC2 • FC3 • FC4 • FC5 • FC6 • FC9 • FC11 • LC7 • LC8 • LC17 • TC10 • TC16 • TC18 • XC13 • XC14 • XC15 • XC19 • XC20	• Flow controller (feed D) • Flow controller (feed E) • Flow controller (feed A) • Flow controller (feed C) • Flow controller (compressor by-pass) • Flow controller (purge) • Flow controller (steam)] • Flow controller (condenser CWR) • Level controller (vapour/liquid separator) • Level controller (separator) • Level controller (reactor) • Temperature controller (reactor CWR) • Temperature controller (stripper) • Temperature controller (reactor) • Composition controller (feed A) • Composition controller (feed D) • Composition controller (feed E) • Composition controller (purge B) • Composition controller (product E)	• Slave controller to XC14 • Slave controller to XC15 • Slave controller to XC13 • Slave controller to LC17 • Standalone • Slave controller to XC19 • Slave controller to TC16 • Standalone • Standalone • Standalone • Master controller to FC4 • Slave controller to TC18 • Master controller to FC9 • Master controller to TC10 • Master controller to FC3 • Master controller to FC1 • Master controller to FC2 • Master controller to FC6 • Master controller to TC16
Analyser		
Tag	**Description**	**Remarks**
• XAr • XBr • XCr • XDr • XEr • XFr • XDp • XEp • XFp • XGp • XHp	• Composition of A in feed • Composition of B in feed • Composition of C in feed • Composition of D in feed • Composition of E in feed • Composition of F in feed • Composition of D in product • Composition of E in product • Composition of F in product • Composition of G in product • Composition of H in product	

• XAw • XBw • XCw • XDw • XEw • XFw • XGw • XHw	• Composition of A in purge • Composition of B in purge • Composition of C in purge • Composition of D in purge • Composition of E in purge • Composition of F in purge • Composition of G in purge • Composition of H in purge	
Stream		
Tag	**Description**	**Remarks**
• Stream 1 • Stream 2 • Stream 3 • Stream 4 • Stream 5 • Stream 6 • Stream 7 • Stream 8 • Stream 9 • Stream 10 • Stream 11	• A feed • D feed • E feed • C feed • Stripper overhead • Reactor feed • Reactor product • Recycle • Purge • Separation liquid • Product	

The TE process has been created with intelligent CAD tool (AVEVA P&ID®). Figure 126 shows the TE schematic without instrumentation (PFD) for clarity. The corresponding connectivity matrix, a zoom in list of plant items in the connectivity matrix and reachability matrix of the connectivity matrix are shown in
Figure 127, Figure 128 and Figure 129 respectively.

For the first time, we now have the TE connectivity matrix people have been trying to generate in an automated way for many years. This provides a framework for other analysis and extension of basic information presented in process P&IDs.

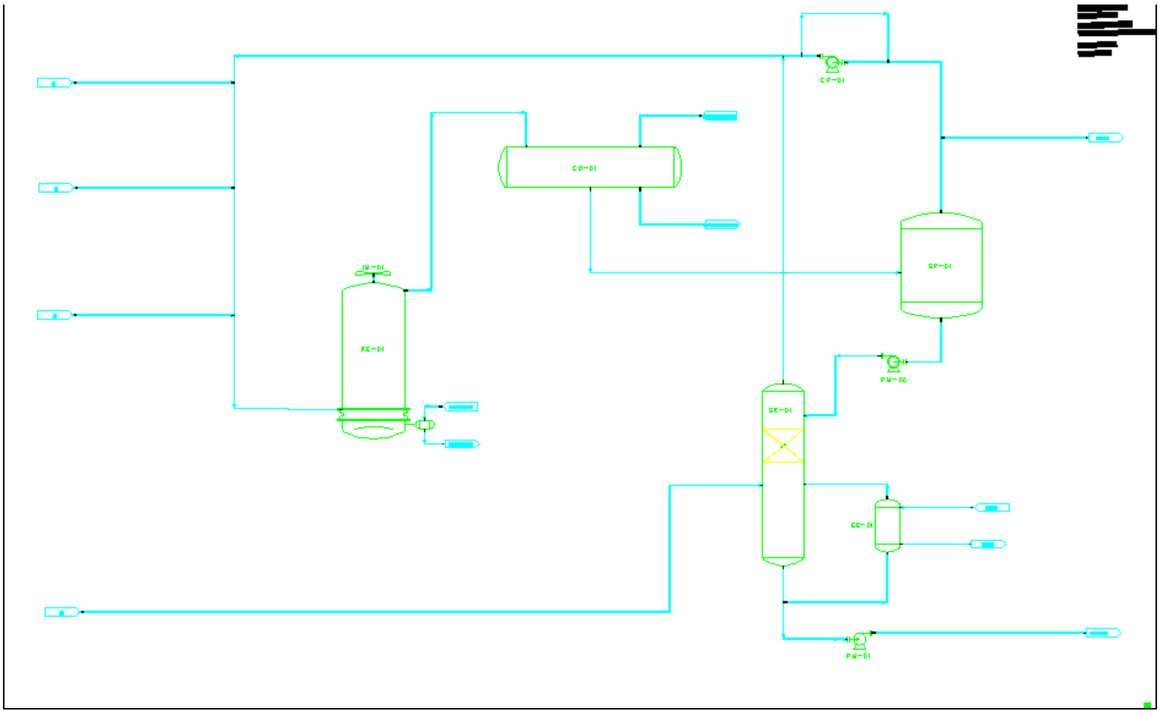


Figure 126: Intelligent CAD drawing of the TE process flow diagram

	C	SR-01	40E-6C3	SP-01	45D-684	45D-B43	CP-01	48F-B37	48F-B38	48F-5E7	48F-4CE	48F-4FF	A	D	E	RX-01	CD-01	PM-02	XE-01	CWS-CC	CWS-RX	SC-01	PM-01	STEAM	PURGE	CWR-CC	CWR-R)	PRODU(	COND.
C	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SR-01	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40E-6C3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0
SP-01	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
45D-684	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
45D-B43	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CP-01	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48F-B37	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48F-B38	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48F-5E7	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48F-4CE	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
48F-4FF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RX-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
CD-01	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
PM-02	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
XE-01	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
CWS-CONDENS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
CWS-RXTOR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
SC-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
PM-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
STEAM	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
PURGE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CWR-CONDENS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CWR-RXTOR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PRODUCT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
COND.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 127: Connectivity matrix of the TE process flow diagram

1		C	SR-01	40E-6C3	SP-01	45D-684	45D-B43	CP-01	48F-E
2	C	0	1	0	0	0	0	0	0
3	SR-01	0	0	1	0	0	0	0	0
4	40E-6C3	0	0	0	0	0	0	0	0
5	SP-01	0	0	0	0	1	0	0	0
6	45D-684	0	0	0	0	0	1	0	0
7	45D-B43	[illegible]	[illegible]	[illegible]	[illegible]	0	0	1	0
8	CP-01	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	1
9	48F-B37	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	0
10	48F-B38	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	0
11	48F-5E7	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	0
12	48F-4CE	0	0	0	0	0	0	0	0
13	48F-4FF	0	0	0	0	0	0	0	0
14	A	0	0	0	0	0	0	0	0
15	D	0	0	0	0	0	0	0	0
16	E	0	0	0	0	0	0	0	0
17	RX-01	0	0	0	0	0	0	0	0
18	CD-01	0	0	0	1	0	0	0	0
19	PM-02	0	1	0	0	0	0	0	0
20	XE-01	0	1	0	0	0	0	0	0
21	CWS-CONDENS	0	0	0	0	0	0	0	0
22	CWS-RXTOR	0	0	0	0	0	0	0	0
23	SC-01	0	0	0	0	0	0	0	0
24	PM-01	0	0	0	0	0	0	0	0
25	STEAM	0	0	0	0	0	0	0	0
26	PURGE	0	0	0	0	0	0	0	0
27	CWR-CONDENS	0	0	0	0	0	0	0	0
28	CWR-RXTOR	0	0	0	0	0	0	0	0
29	PRODUCT	0	0	0	0	0	0	0	0
30	COND.	0	0	0	0	0	0	0	0

Pipe in-line elements such as a pipe tee automatically inserted by intelligent CAD tool

Figure 128: Zoom in row and column elements of TE connectivity matrix

1		C	SR-01	40E-6C3	SP-01	45D-684	45D-B43	CP-01	48F-B37	48F-B38	48F-5E7	48F-4CE	48F-4FF	A	D	E	RX-01	CD-01
2	C	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
3	SR-01	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
4	40E-6C3	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
5	SP-01	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
6	45D-684	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
7	45D-B43	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
8	CP-01	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
9	48F-B37	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
10	48F-B38	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
11	48F-5E7	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
12	48F-4CE	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
13	48F-4FF	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
14	A	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
15	D	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
16	E	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
17	RX-01	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
18	CD-01	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
19	PM-02	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
20	XE-01	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
21	CWS-CONDENSER	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
22	CWS-RXTOR	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
23	SC-01	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
24	PM-01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	STEAM	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1
26	PURGE	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	CWR-CONDENSER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	CWR-RXTOR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	PRODUCT	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	COND.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 129: Reachability matrix of the TE process (PFD). Entries marked red indicate elements involved in a recycle via process fluid or signal flow while green colour signifies reachability from row elements

The intelligent drawing forms the basis and starting point for the TE connectivity analysis. The ISO15926 compliant XML files exported from such intelligent CADs are parsed to extract relevant plant items and to generate the connectivity matrix which can be manipulated for plant analysis. For example, the reachability matrix shown in Figure 129 is produced from the process connectivity matrix and suggests that any disturbance emanating from the steam supply to vapour liquid stripper has the potential to affect all plant items except the Column plant items that are not covered in green colour i.e. streams A, D, E, Condenser CWS, Reactor CWS, and SC-02.

The entries of the connectivity matrix in red colour are the non-zero diagonal elements of the reachability matrix. They indicate plant items involved in a recycle network. For example, SR-01, SP-01 and CP-01 are part of the process recycle units while PM-01 is not associated with the process recycle system.

The illustrated example shows that the reachability matrix of a process plant can be utilized to provide a high level insight into the underlying process through its connectivity and directionality information.

The next demonstration of the TE analysis considers the process with implemented control structure, control loops and instruments. Process schematic with implemented control scheme of the TE to be used as the basis for hypothesis testing in this report is shown in Figure 130. The control scheme is taken from the work reported by Chiang and Braatz (2003).

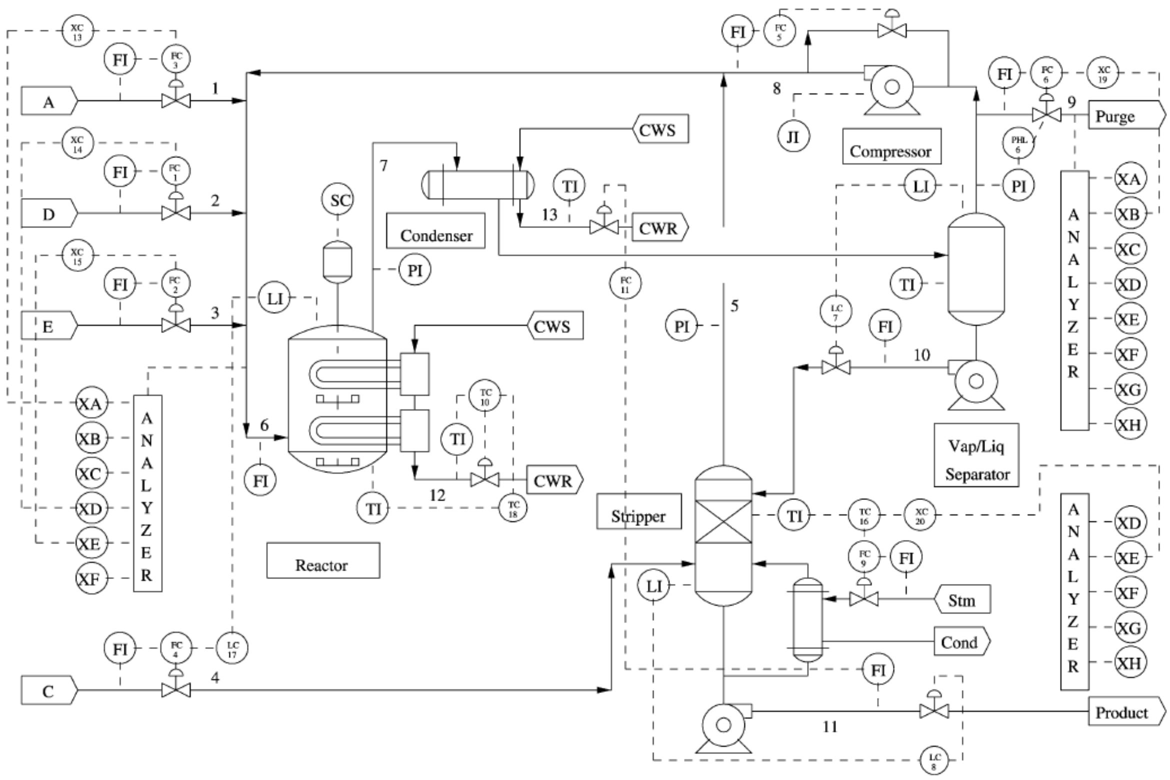


Figure 130: TE process schematic according to Chiang and Braatz (2003) with Lyman and Geogakis (1995) plant-wide base control structure

An electronic intelligent P&ID drawing of the TE process using the tags listed in Table 15 is depicted in Figure 131. The ISO15926 XML compliant of the intelligent P&ID is parsed and manipulated by the connectivity tool to generate the connectivity matrix shown in Figure 132.

With the parsing of the TE XML and creation of connectivity matrix, the process connectivity tool is ready to carry out automated analysis to test some of the hypothesis proposed in Chiang and Braatz (2003) about the feasible propagation paths.

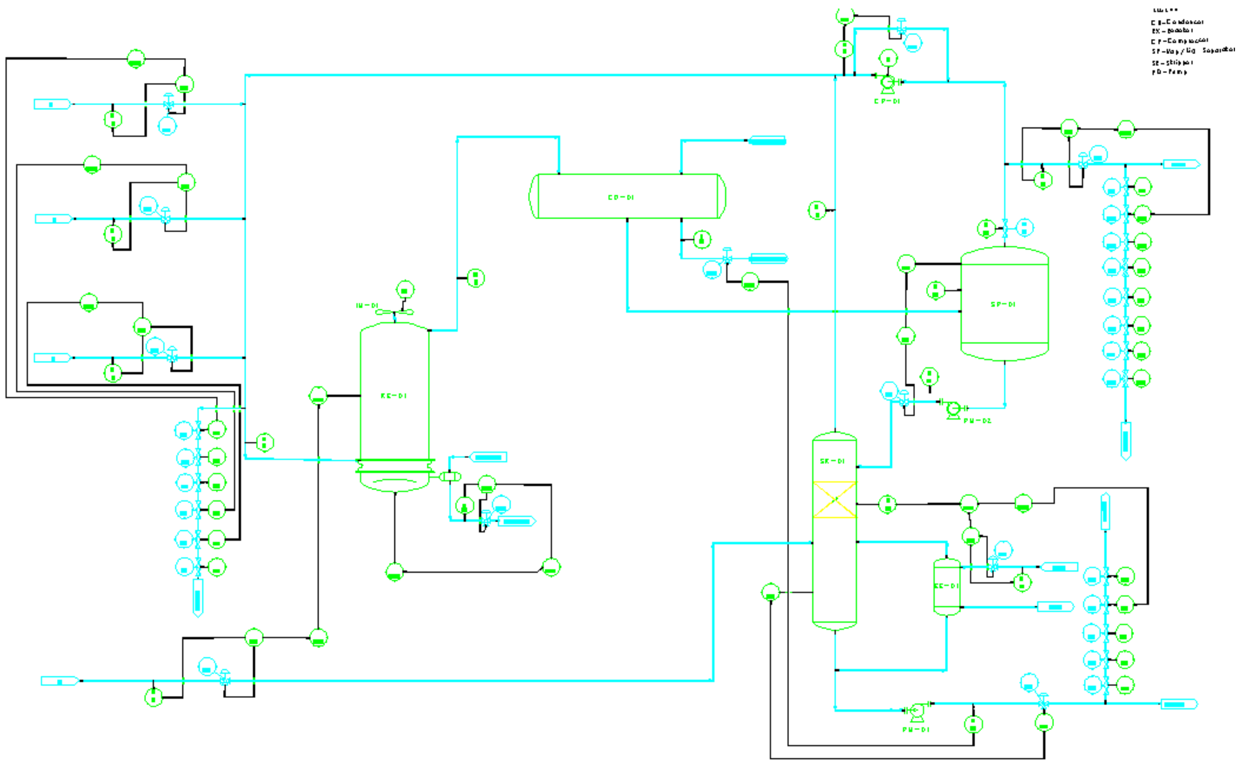


Figure 131: Intelligent CAD drawing of TE using AVEVA P&ID® and AutoCAD®

5.7.2 Data – driven Analysis

The multivariable data-driven methods described by Chiang and Braatz (2003) are called *modified distance* and *modified causal dependency.* They are used for identifying broken sensors and broken causal dependencies respectively. The traditional *distance* utilizes the historical data of process sensor measurements to estimate changes in the frequency distribution of process while the *causal distance* is based on the relationship between the frequency distributions of two variables. A residual greater than a preset threshold signifies a broken sensor or broken causal dependency and thus a fault.

The main contribution of the modified approaches reported by Chiang and Braatz (2003) is the use of continuous-time model as opposed to the use of bins to group measurement data and define distribution with the attendant loss of resolution depending on the size of the bin.

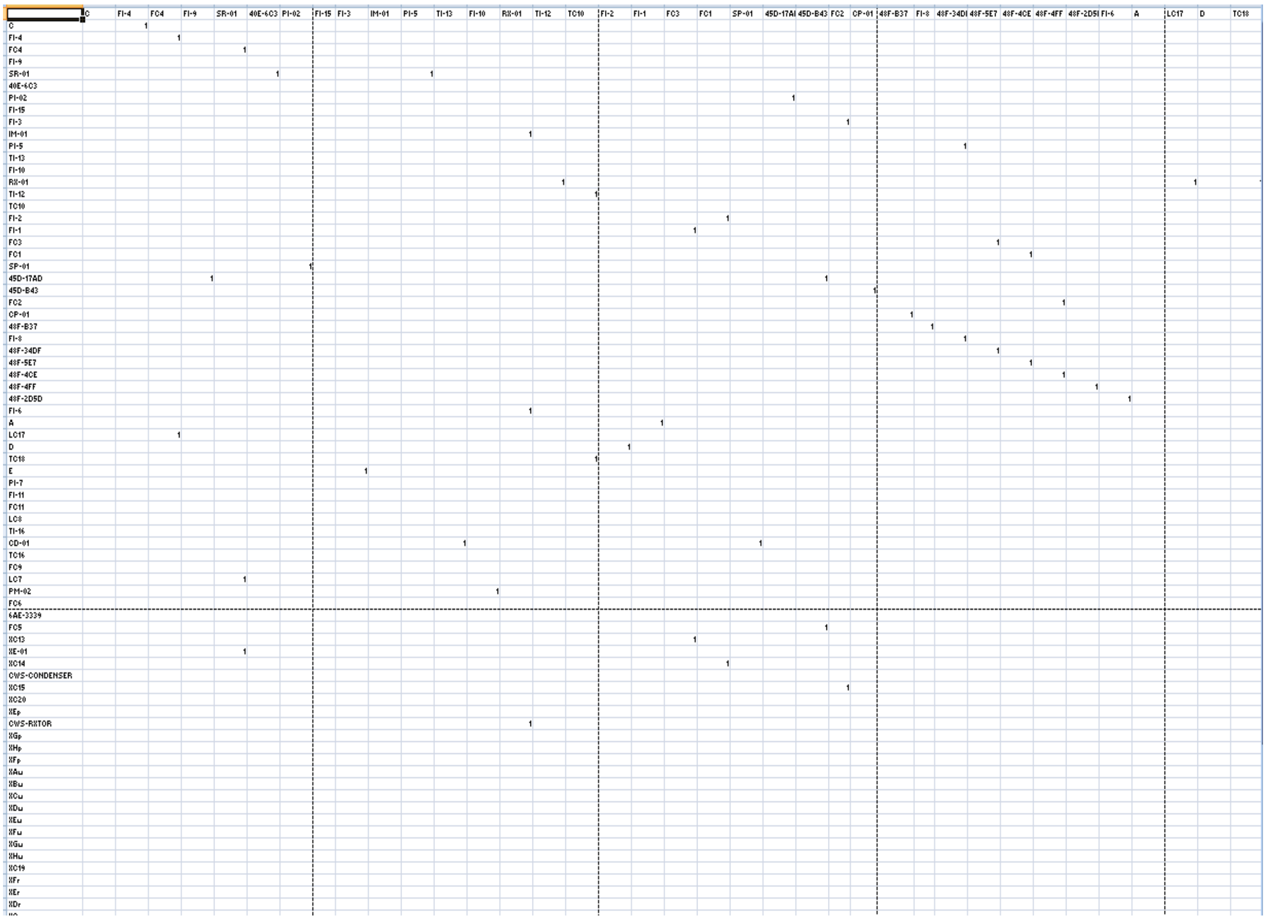


Figure 132: A portion of connectivity matrix of the TE P&ID

5.7.3 Hypothesis Testing using Process Connectivity Analyser

For the testing and evaluation of the methods described in Chiang and Braatz (2003) a subset of process faults listed in Table 16 is considered. The authors examined the process schematic and manually constructed the following fault propagation paths originating from FC3:

Propagation from FC3 ultimately to XAw (purge)
Propagation from FC3 ultimately to CP-01 (compressor)
Propagation from FC3 ultimately to XAr (feed)

Table 16Results from data-driven analysis for fault 6 (A feed loss in stream 1) identified the feed valve for reactant A (FC3) as the root-cause for fault 6. Chiang and Braatz (2003) also observed that as time progressed, the fault affected other variables downstream.

The authors examined the process schematic and manually constructed the following fault propagation paths originating from FC3:

- Propagation from FC3 ultimately to XAw (purge)
- Propagation from FC3 ultimately to CP-01 (compressor)
- Propagation from FC3 ultimately to XAr (feed)

Table 16: TE process faults

Variable number	Process variable Description	Type
IDV(1)	A/C feed ration, B composition constant (Stream 4)	Step
IDV(2)	B composition A/C ratio constant (Stream 4)	Step
IDV(3)	D feed temperature (Stream 2)	Step
IDV(4)	Reactor cooling water inlet temperature	Step
IDV(5)	Condenser cooling water inlet temperature	Step
IDV(6)	A feed loss (Stream 1)	Step
IDV(7)	C header pressure loss-Reduced availability (Stream 4)	Step
IDV(8)	A, B, C feed composition (Stream 4)	Random variation
IDV(9)	D feed temperature (Stream 2)	Random variation
IDV(10)	C feed temperature (Stream 4)	Random variation
IDV(11)	Reactor cooling water inlet temperature	Random variation
IDV(12)	Condenser cooling water inlet temperature	Random variation
IDV(13)	Reaction kinetics	Slow drift
IDV(14)	Reactor cooling water valve	Slow drift
IDV(15)	Condenser cooling water valve	Sticking
IDV(16)	Unknown	
IDV(17)	Unknown	
IDV(18)	Unknown	
IDV(19)	Unknown	
IDV(20)	Unknown	
IDV(21)	The valve for stream 4 was fixed at the steady state position	Constant position

The process connectivity tool would have aided in the Chiang and Braatz (2003) in their analysis and automate testing for feasible fault propagation of the above hypothesis from data-driven analysis without recourse to the manual path tracing.

The analysis with *Process Connectivity Analyser* starts with a summary of tagged items in the TE plant as shown in Figure 133. The figure shows parsed tags directly from the electronic XML description of the TE plant.

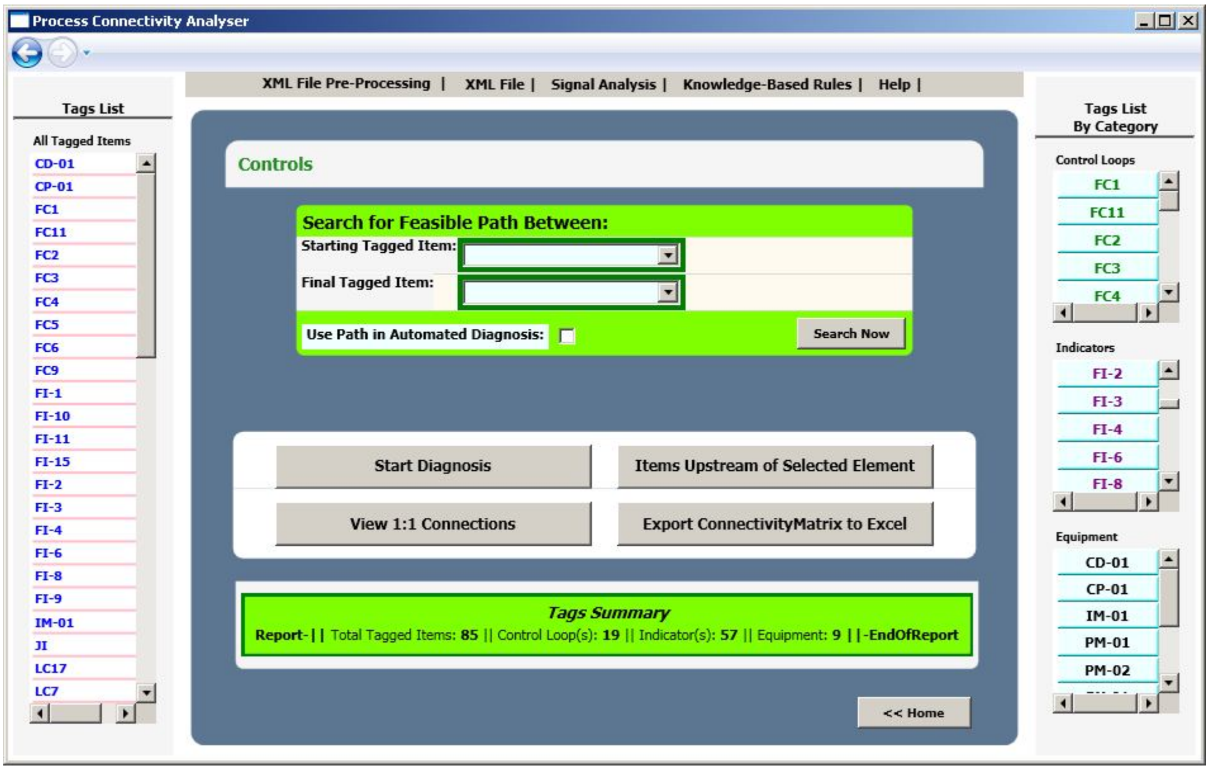


Figure 133: Parsed plant items from the TE process

In the first instance, the connectivity tool would have helped to identify all the plant items/measurement points that would be affected by the suspected data-driven root-cause by looking at plant items downstream of FC3 (Section 5.3.1 functionality *L* of Figure 66 of the thesis).

Secondly to confirm the existence of a feasible propagation path between a suspected root-cause and the final secondary disturbed tag. This is demonstrated in Figure 134, Figure 135, Figure 136 and Figure 137.

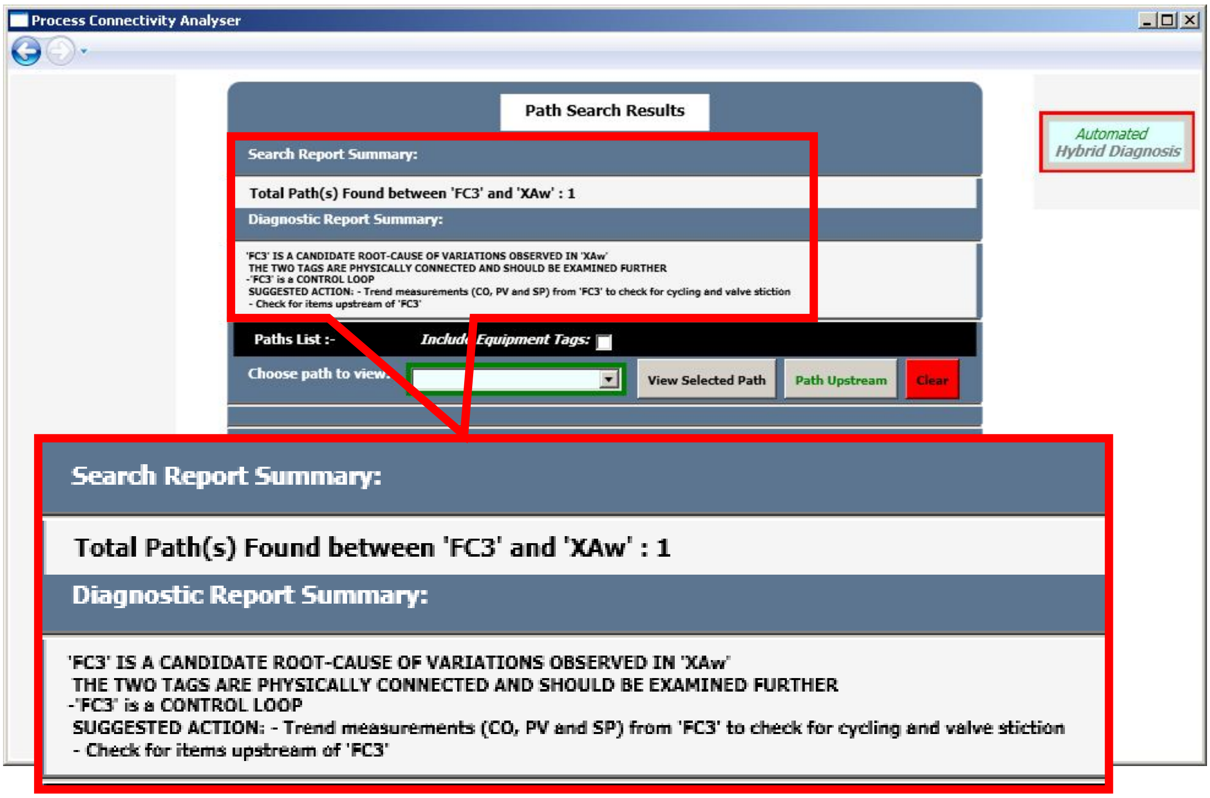


Figure 134: Process Connectivity Analyser confirmation of propagation from FC3 to XA in purge due to fault 6

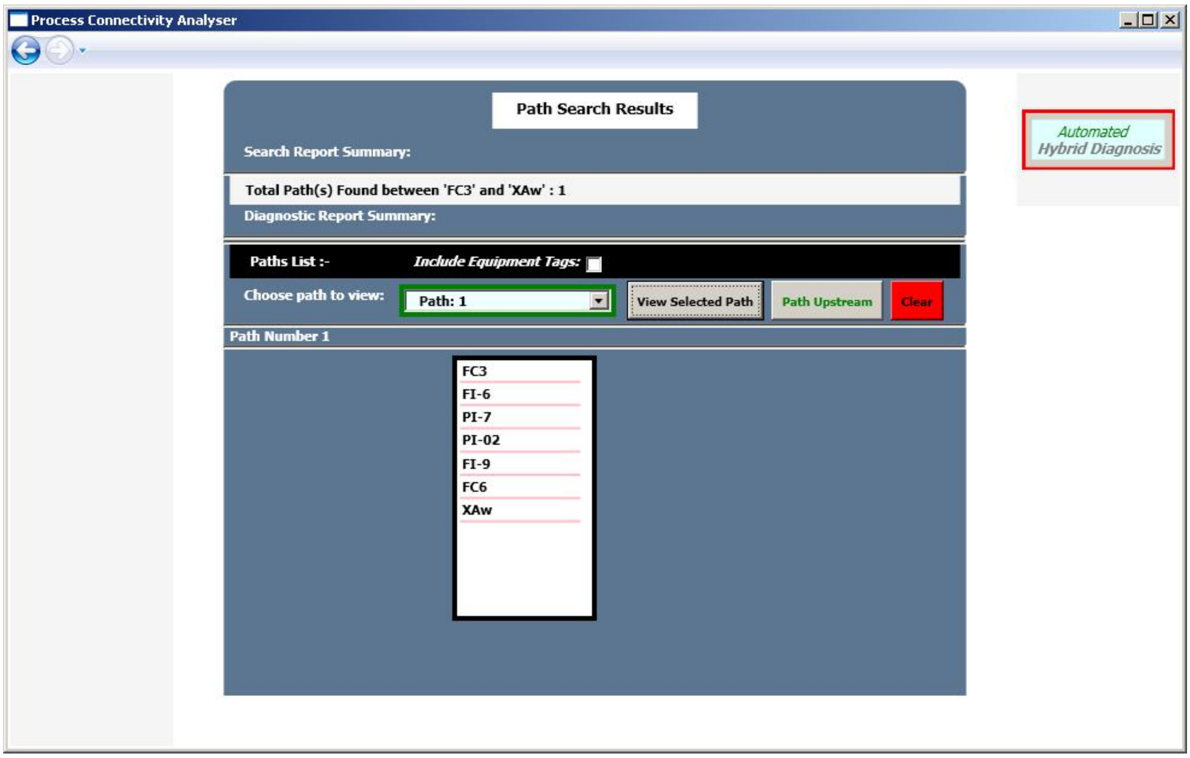


Figure 135: Confirmation and list of plant items in the forward path of fault propagation to XA in feed due to fault 6 (measurement points and control loops)

The connectivity tool is also able to include processing units with the measurement points as well as the control loops. This capability is lacking in pure data-driven tools as they only consider measurement points. However, some faults are actually caused by processing equipments and picked up at measurement points, in some cases some distance away from the source (e.g. the use of proxy measurement in Thornhill, *et al.*, (2003)). With pure data-driven tools, process control engineers have to resort to electronic or paper print out of process schematic to identify the root-cause. Experienced control engineers would make use of their mental model of the process. The process connectivity tool eliminates the need for this and therefore opens up process fault propagation analysis to a wider range of users.

The remaining analysis of fault 6 in the TE process will incorporate processing equipment with measurement points and control loops in listing tags in fault propagation paths.

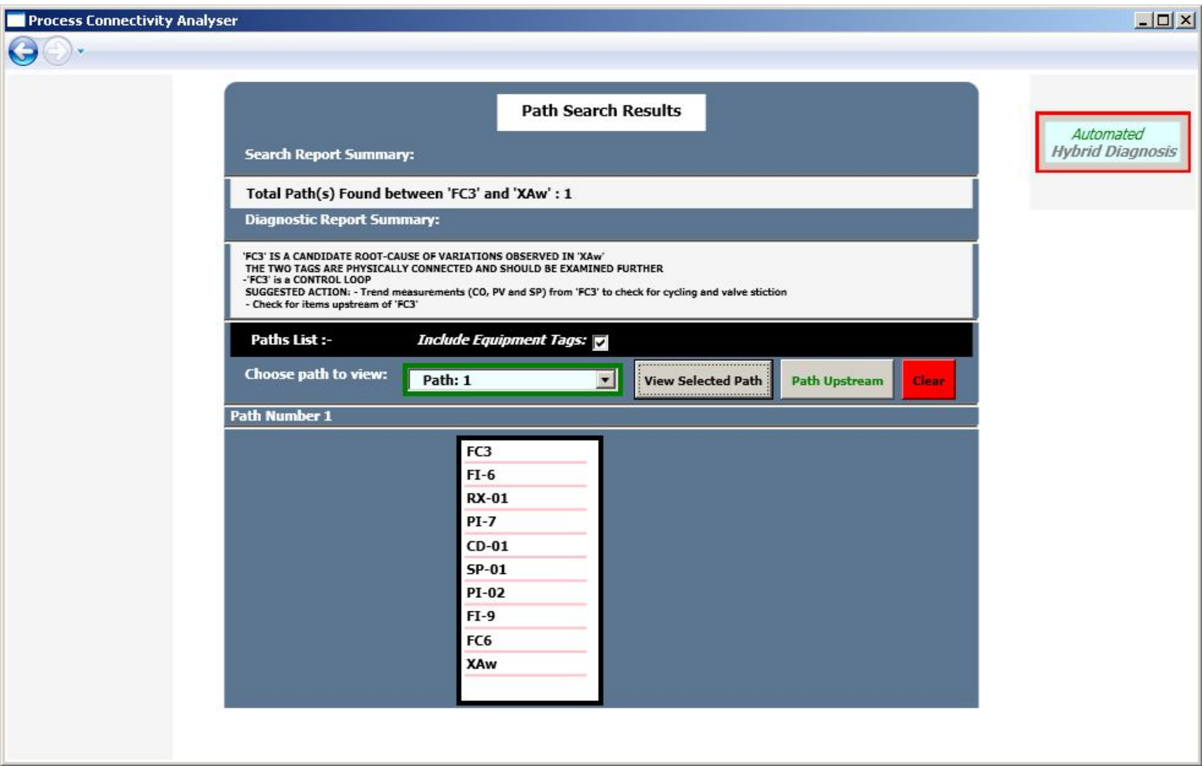


Figure 136: Confirmation and list of plant items in the forward path of fault propagation to XA in purge due to fault 6

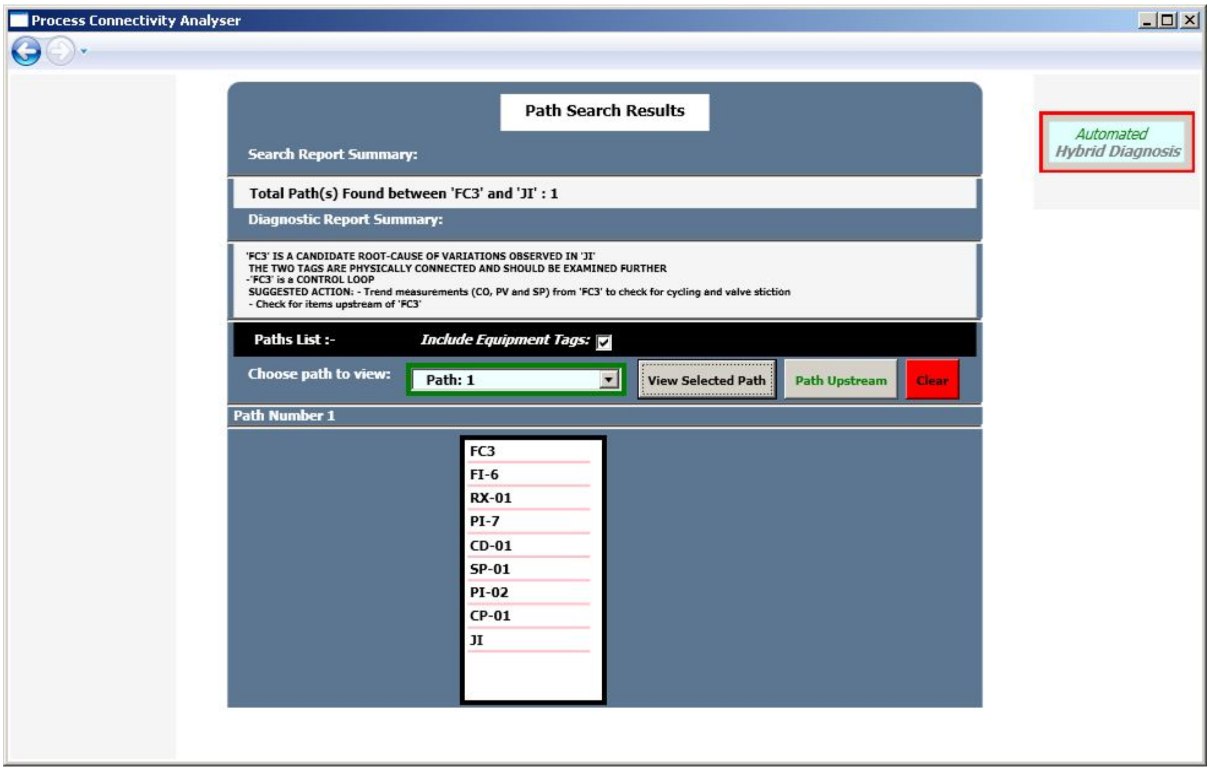


Figure 137: Confirmation and list of plant items in the forward path of fault propagation to recycle compressor due to fault 6

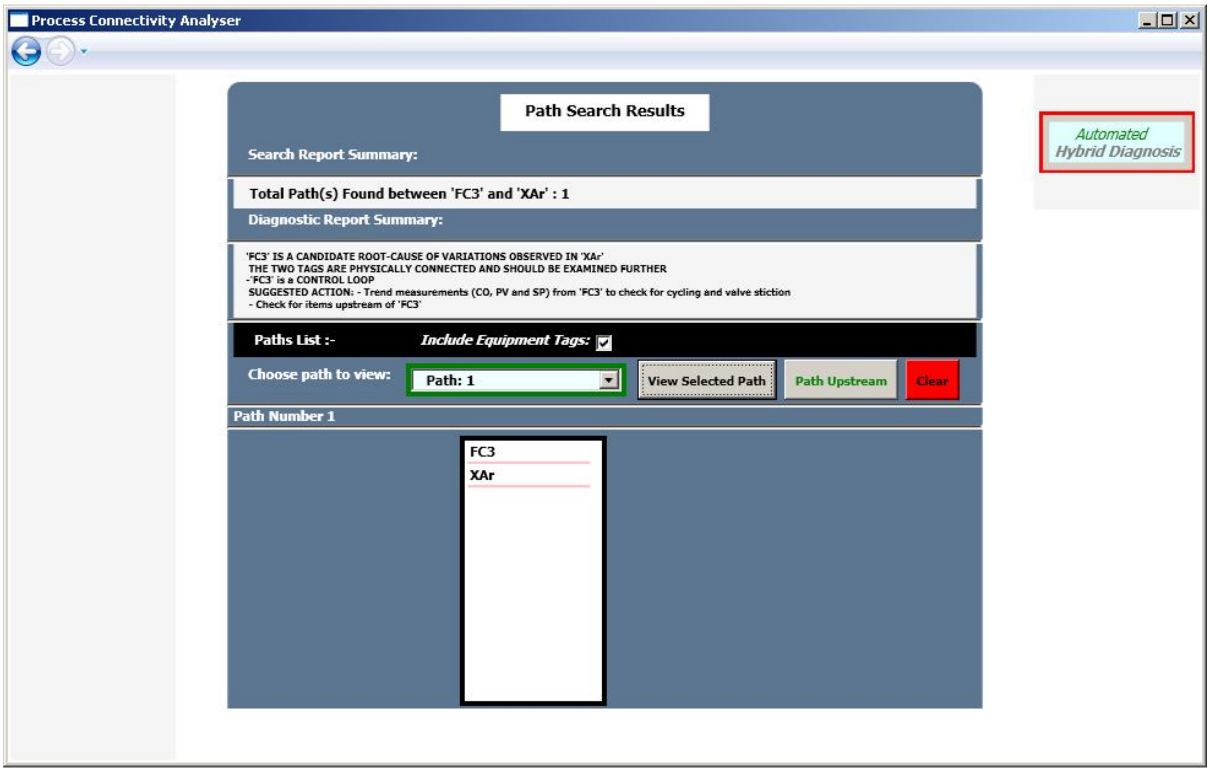


Figure 138: Confirmation and list of plant items in the forward path of fault propagation to XA in feed due to fault 6

5.8 Chapter Summary

Chapter five has presented an overview and operational procedures of *Process Connectivity Tool,* an output of the PhD work. To demonstrate the effectiveness and versatility of the tool developed, an illustrative example, an industrial case study and an academic case study were considered. The results obtained suggested that the tool can be used as a standalone for process analysis, in` combination with results from data-driven analysis and for hypothesis testing of results from data-driven methods.

6 Industrial Discussions and Context

This chapter relates industrial contacts made during the course of this work. It discusses the contacts made with representatives from commercial software vendors relevant to the project. The chapter also focuses on commercial tools (e.g. SmartPlant® from Intergraph® Bentley AutoPlant P&ID® from Bentley®, Aveva P&ID® from Aveva®, AutoCAD P&ID® from AutoDesk®) suitable for electronic, automated capture of complex chemical and petrochemical plants such as a refinery on an industrial scale. Commercial tools used in the process industries to generate key performance indicators (KPI), a quantifiable metric indicating how well a plant is performing and for diagnosing malfunctions, using process measured data such as temperature, pressure, level and so on are also considered.

Consideration of KPIs is important because an improved diagnostic is achieved when KPIs from data-driven diagnosis tools are combined with connectivity information contained in process plant representation such as in a P&ID and process know-how (Thambirajah, 2008; Scherf *et al*., 2006 and Yim & Ananthakumar, 2005).

6.1 Result Formats from Data-Driven Tools for use with Connectivity Tool

Many signal processing and analysis tools are available in the market to extract useful information from enormous quantities of measurement data generated on a regular basis from process plants. One example of a commercial tool is the PlantTriage® from ExperTune®, and those listed in Table 17. Most of these tools acquire real time data from plant's DCS to continuously monitor the state of the plant. Several key performance assessments, discussed in Section 6.1.1, are calculated for each control loop and used to monitor several controller properties. Each assessment is calculated at some specified periods and plant operator can select which performance metrics are to be used to create the loop health assessment which gives an indication of the overall health of the control loop and ultimately health of the plant.

Process control engineers diagnose malfunctions by examining and comparing the chosen key performance indices (KPIs) with the best achievable standard already in place and request corrective action on the loops with a poor performance.

However, any conclusion reached from data-driven analysis has to be viewed against the process topology to ensure that the results from data analysis make sense. The procedures for reaching conclusion when KPIs generated by signal processing tools are tested against process topology is automated in this thesis by utilizing process connectivity and directionality information derived from process schematics such as a P&ID.

6.1.1 Key Performance Indicators (KPIs)

This section discusses the relevance of KPIs, in an industrial setting to the thesis. It also provides a list of major commercial vendors of data-driven software tools for generating KPIs for monitoring and diagnostic purposes.

A wide range of numerical performance indicators values such as control variable average error, oscillation index, set-point crossing, valve movement and valve stiction are used in the process industries such as refineries to measure individual control loop performance and overall performance of the plant when compared to the ideal or best performance possible. The comparison ensures delivery of the control and monitoring strategic objectives such as reduction in product variability and plant-wide oscillation (Harris, *et al*., 1999; Horch and Isaksson, 1999; Jelali, 2006; Thornhill, *et al*., 1998).

Most commercial data-driven tools such as PlantTriage® from ExperTune® for generating KPIs connect to plant's distributed control system (DCS) to acquire and process real time data to generate KPIs for checking performance levels. Major commercial vendors listed in Table 17 take process measurements, usually in real time, to generate important key performance indices/indicators such as average absolute error (AAE), Harris index, output standard deviation (OSD), oscillation index, variance, process lag, process gain, set point crossings, variability, and valve stiction which are used by control engineers to diagnose anomalies and optimize plant performance.

In this thesis, the approach is to combine some of the KPIs with process connectivity information and utilize process know-how to draw conclusion about the root-cause of a plant-wide disturbance.

Table 17: Commercial software tools for measured signal processing and analysis.

Software	Vendor
PlantTriage®	ExperTune®
Loop Performance Manager®	ABB®
LoopScout®	Honeywell®
Control Performance Monitor®	Matrikon® (owned by Honeywell®)
Loop Analysis® (formerly Control Wizard®)	PAS®
PID Watch®	Aspentech®
PDA®	ABB®
Performance Watch®	Invensys®
Control Monitior®	Control Arts®
PCT Loop Optimizer®	ProControl Technology®
PROBEwatch®	ISC®
INTUNE+	ControlSoft
Control Loop Performance	Capstone Technology
Plant ESP	Control Station
DeltaV Insight	Emerson
rCAAM (RoviSys Control Assessment and Monitoring)	RoviSys

6.1.2 Section Summary

The section described leading commercial software packages available for data analysis.
PDA tool from ABB has been chosen for data-driven root-cause analysis in this thesis because Imperial College has a licence in place. However, the connectivity tool can also be combined with other commercial data-driven analysis tool to provide insight into the mechanisms of disturbance propagation and to make sense of the results from pure data analysis.

6.2 Meetings with Representatives from Software Companies

Meetings were arranged and held with representatives from commercial software tools, notably some of those listed in Table 17 to discuss what capabilities their software can offer. Most discussions were held during a software evaluation and selection period. Specifically, meetings were held with representatives from Intergraph, Hazid® Technologies Limited and Aveva. Hazid® Technologies Limited has some experience in using some of the tools that are of interest to the research. The meetings yielded positive contributions to the project and are well documented. In some cases, emails were exchanged for the purpose of sharing ideas, knowledge and experience.

The author also liaised with engineers at AVEVA having chosen the company's software (AVEVA P&ID) for the project with the collaboration leading to identification and fixing of bugs in the export of ISO15926 compliant XML. For example, some drawing items were not included in the XML export which was reported back to AVEVA representative with a view to fix the bug.

6.3 Industrial Placements at Sponsor's Sites

This section discusses some of the experiences and lessons learnt while on industrial placements at the research sponsor's sites. There were effectively two work experiences during the course of the research work.

The first placement was based at a technology support centre where processes on sites are monitored remotely via web-based data-driven monitoring tools. The monitoring tools use process measurement data for analysis. Key performance indicators (KPIs) are generated based on the measurement analysis. The KPI is a matrix that gives an indication of the state of the process health and help process control engineers to identify poorly performing loops.

In reaching conclusions before major maintenance efforts are initiated, control engineers typically reconcile results from data-driven tools with the mental physical topology of the plant under consideration and in some cases using the actual process schematic laid out. This is the crux of this thesis-research into innovative ways to capture process connectivity and directionality information from readily available sources to enhance and provide insight into pure data-driven analysis. The placement predominantly involved preliminary studies on the techniques in place for identifying root-causes with a view to investigate how the analysis might be combine with the proposed connectivity tool to be developed in the research.

The second placement consisted of onsite and remote monitoring operations. Real process plant was analysed and drawn with intelligent CAD tool to generate ISO15926 compliant XML as well process connectivity matrix from the connectivity analyser tool developed. Based on the discussions held with onsite process control engineers, the following practical considerations came to the fore:

- Disturbance propagation is not always one way, even when this is not indicated on the process P&ID
- Multiple root-causes may be accounted for
- Disturbances other than oscillations do affect process plant performance
- High frequency disturbances may not be detected from practical data capturing sampling rate and storage perspective
- Some process lines on the P&ID may not be in use. It will be useful to indicate this in an automated way with a software tool.

The bottom line of the observations above is that not everything needed for effective and complete process analysis is captured by the process P&ID. However, the advantage of using process P&IDs stem from the fact that they are readily available in industries at no extra cost and efforts, just like historical data are readily available for signal analysis.

The remarks also show that there is huge potential for future research opportunities as the process P&IDs can be augmented either at the drafting stage or after XML export.

6.4 Chapter Summary

This chapter has presented a summary of contacts made with commercial software vendors and project sponsor during the course of the research. The chapter reinforces the multi disciplinary nature of the thesis and the approach adopted in the collaborative efforts to ensure a successful and timely completion of the research.

7 Summary and Opportunities for Future Work

This chapter details the summary of work reported in this thesis and possible future research opportunities with suggested strategy for implementation. The summary section reviews important activities reported in the thesis that resulted in the accomplishment of the thesis aims and objectives. The other section of this chapter proffers suggestions on future work with respect to further research opportunities and improvement to existing implemented methodologies.

7.1 Summary

The report has described the activities and approaches taken to accomplish research aims and objectives set out in the thesis. This section considers each key task from the various chapters discussed in the thesis. It started with discussions on the background and emphasizes the motivation for carrying out the project. The section closes with major conclusions from the thesis.

7.1.1 Aims and Objectives of the Thesis

Key objective of process control strategy is to divert process variability away from key process variables to places that can accommodate the variability. Modern process plants such as refineries implement advanced process control systems, reduced inventory, heat integration, recycle streams , and back-up and recovery system in other to meet business objectives resulting in highly coupled plant with strong process dynamic interaction. The resulting complex plant complicates root-cause diagnosis of disturbances because a local process upset propagates to units downstream creating secondary plant-wide effects.

The motivation for the work reported in this thesis is given in Thornhill *et al*., (2003). The paper described a data-driven root-cause analysis that found a sticking valve in a chemical plant. It concluded that an engineer needs to combine knowledge from the process schematic with the results from data-driven analysis in order to complete the analysis.

The aim of the research therefore is to conduct research into and establish new ways to capture and manipulate information from process schematic in order to give an automated means of diagnosing plant-wide performance problems. The project utilized electronic representation of process schematic using XML to automatically generate process connectivity matrix. The practical outcome will be improved disturbance diagnostics using process data, information from a process schematic and process understanding (know-how).

The project concerns automated capture of connectivity information in large plants such as refineries, the linkage of this connectivity information with the results from data-driven cause-and-effect analysis of the process measurements, and incorporation of process know-how to draw conclusions about the causes of disturbances. *Connectivity information* means a specification of the items in the plant and the connections between them in a form that can be manipulated algorithmically using a computer program. An example of process know-how is the existence and mechanism of the destabilizing effect that heat integration can have on a process.

Automated tools for tracking down the root-cause of plant-wide disturbances are essential in modern complex, interrelated and highly coupled process plants because a plant running smoothly with little or no disturbance makes the most profit. Large scale industrial requirements for plant-wide approach to process analysis, control, diagnosis and optimization have been identified. These requirements include:

- Facility-wide benchmarking and standardization of control systems;
- Characterization of performance faults;
- Detection of the presence of one or more periodic oscillations;
- Detection of non-periodic disturbances and plant upsets;
- Determination of the locations of the various oscillations/disturbances in the plant and their most likely root-causes;
- Incorporation of process knowledge such as the role of each controller;
- Automated model-free causal analysis to find the most likely root-causes.

The objectives set out at the beginning of the thesis have been met. A software tool that demonstrates large scale electronic capture of process connectivity and directionality information in an automated manner has been designed and developed during the course of the PhD work. The software tool developed was used to combine connectivity information with signal analysis and process know-how to test and draw a conclusion about hypotheses from process signal analysis. Other uses such as finding the location and nature of a plant item on the process P&ID, checking for full connectivity at the drawing stage and export of full process connectivity matrix to Excel application have also been demonstrated.

A review of prior work led to appreciation of the various approaches taken by several researchers to address the requirements listed above and to put the PhD work in context.

The following conference and journal papers have been produced from the research work:

- Di Geronimo Gil, G.J., Alabi, D.B., Iyun, O.E. and Thornhill, N.F., 2011, Merging process models and plant topology, *Advanced Control of Industrial Processes* (*ADCONIP* 2011), Hangzhou, China, May 23-26 2011.
- A journal paper on the Tennessee Eastman (in progress).

This is in addition to technical reports, several presentation slides and monthly meetings minutes delivered to the project sponsor during the course of the research work.

7.1.2 Process Representation

The project utilized electronic representation of a process schematic using ISO15926 compliant XML. XML is a platform and vendor independent approach to data storage and transmission which is fast becoming the *de facto* standard in the information technology industry (Girardot and Sundaresan, 2011).

Signed digraphs and multi flow modelling (MFM) are competing technologies to the use of XML. XML was chosen as the technology of choice in this project for process schematic data encoding standard because of its flexibility and acceptability within the leading commercial CAD tool vendors. For example AVEVA has incorporated XMpLant translation engine from Noumenon Consulting for ISO15926 implementation from proprietary data formats in its intelligent CAD tool called AVEVA P&ID. This meets some of the project requirements for automated conversion and legacy drawing issues of process schematics to electronic format such as XML and also in dealing with legacy drawings.

Process schematics with additional information behind the graphic display such as been utilized in this work to export XML description of process topology are said to be *intelligent* otherwise they are referred to as *dumb*. Intelligent P&IDs used throughout this thesis were created with AVEVA P&ID software tool running on AutoCAD tool. The procedures for transforming a process schematic such as a process flow diagram to electronic XML representation and connectivity matrix is summarized and depicted graphically in Figure 139. XML description of the process topology and results from data-driven analysis are fed into the *Process Connectivity Analyser* tool.

Due to the flexibility and ease with which XML files are created, rules otherwise known as schemas are enforced for valid XML files. Prominent schemas for XML structure for process plant schematics are:

- CAEX based on IEC/PAS 62424, and
- XMpLant based on ISO15926

The two standards for XML encoding were discussed in the thesis. The fact that ISO15926 has been widely adopted and implemented by leading CAD vendors as part of the project requirements was chosen.

XML Parsing and Representing Connectivity

The ISO15926 compliant XML of process schematic is read and parsed by parsing algorithm of the connectivity tool using *XElement* class of Microsoft .NET Language Integrated Query to XML (XLINQ) namespace for reading and manipulation of in memory XML tree data structure to generate process connectivity matrix. The connectivity matrices indicate how plant items are connected to one another. The connectivity matrix generator component of the connectivity tool creates process schematic connectivity matrix with an option to export the connectivity matrix to Excel application.

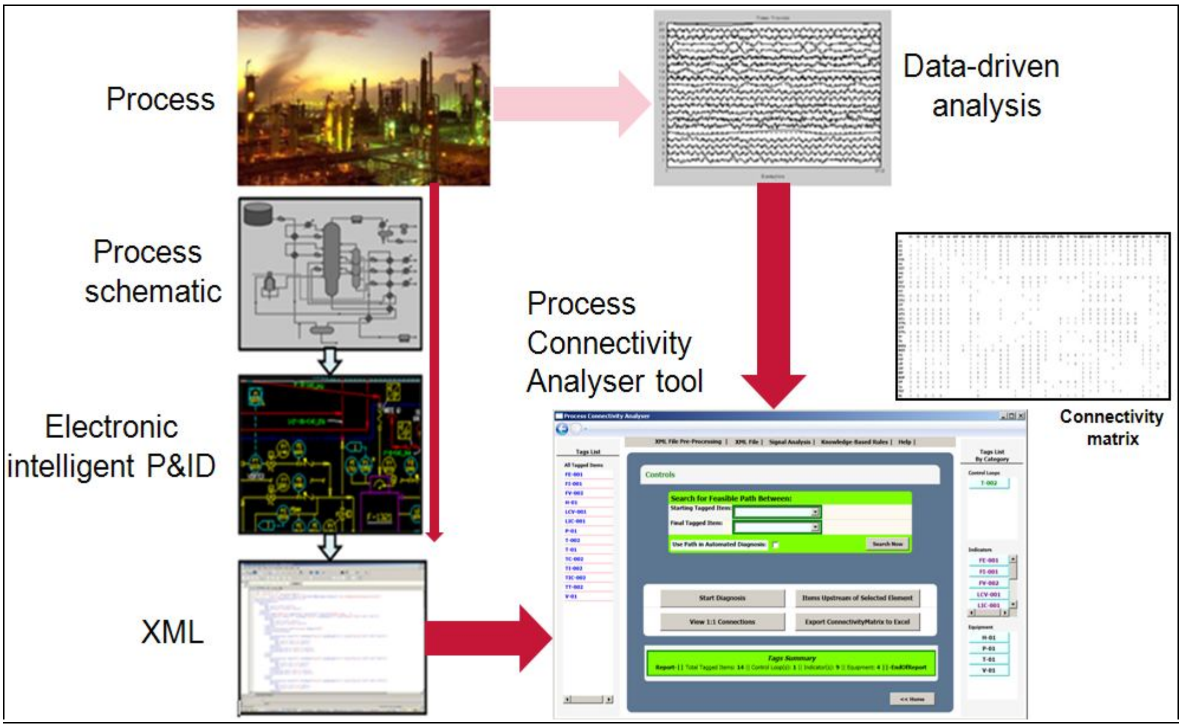


Figure 139: Inputs to *Process Connectivity Analyser* and schematic representation of processes involved in transforming process schematic to intelligent P&ID, electronic XML description and connectivity matrix

Integration of Process Connectivity Information with Results from Data-Driven Analysis and Process Know-how

Process Connectivity Analyser, the final integrated tool, combines process topology representation as connectivity matrix with results from the cause-and-effect data-driven analysis with process understanding. The tool provides a means for manipulation of data, process connectivity information and engineering know-how. Process connectivity information is the form of an XML description of a process flow sheet prepared in an industry standard format. Data is the measurements from the process. Engineering information was made generic and can be customized to meet specific business need such as in the form of rules.

The software implementation is in Microsoft .NET platform and allows users to upload process XML description and results from data-driven analysis from a storage device and the user can perform analysis. Process know how is in built and used in displaying final results for analysis.

Requirements definition for the software were iterative throughout the project as they were regularly updated based on outcomes of regular meetings over the phone and in person throughout the duration of the project. A familiar navigational window-based graphical user interface was designed and implemented to enhance the tool's usability and minimize learning curve by the users.

7.2 Illustrative Example and Case Studies

The illustrative example of the crude heating unit was included to demonstrate the capability of the software tool in a simple and easy to understand manner. The purpose of the industrial case study discussed was to validate research findings by testing research methodology on real life process plants available at the sponsor's site with some modification to the naming conventions to protect confidential information and trade secrets. Nevertheless, the complexity of the case study is exactly as that of the real life plant. The case study demonstrated that connectivity tool is capable of identifying spurious results from pure data-driven root-cause analysis and perform other tasks in an automated and systematic way. The academic case study presented the analysis of a process schematic that is well known and researched in the academic community. This will enable readers to appreciate the capabilities of the connectivity tool since most would have come across the process under consideration and thus have a mental model or some expectations from the tool's output.

7.3 Alternative Uses

This section discusses other uses to which the software tool developed in this work can be put to. This is in addition to the use in root-cause diagnosis of plant-wide disturbance.

- *Model Development in Multivariable Analysis*: The tool developed in this work can be used to check models in advanced process control and multivariable analysis. An example is model development for model in Model Predictive Control (MPC) when dynamic models are to identified from plant test data. The model builder would like to check for unobservable and weakly observable response form a perturbed variable called the *handle*. By using the connectivity tool to find all plant items physically connected to the *handle,* the model builder is able to separate spurious response and carry out further examination on variable that are physically connected to the *handle* but not observable from the perturbation test.

- *Alarm Management*: Process Connectivity Analyser can be used to minimize alarm overload and/or eliminate false alarms so that process operator can respond to and deal with real alarm by finding physical connections among alarms and isolating extraneous alarm.
- *Hazard and Operability Analysis*: The connectivity tool can be used to study hazard and operability analysis of process plant by performing cause-and-effect analysis and path tracing of process fluid and signal flow around the plant. If an area within the plant needs to be shot down and isolated for health and safety reasons, the connectivity tool can be used to identify which part of the production plant will be affected.
- *Sensor Location*: The connectivity tool can be used to identify spots within the plant where sensor(s) should be installed relative to a measurement point. This will eliminate the use of proxy measurement (Thornhill, *et al*., 2003) which will require knowledge of the process plant and more efforts to locate.

7.4 Future Research Ideas and Discussion

This section discusses avenues for improving the methods described in this thesis and future research opportunities. The objective is to make recommendations for continued research on enhancing electronic capture and use of process connectivity information for better insight into plant-wide performance issues.

7.4.1 Implementation Alternatives

This section enumerates alternative approaches for implementing some components of the connectivity tool. This will take the software tool beyond academic prototype and full commercial potential can be unleashed.

Expert System Shell for Knowledge Representation and Reasoning

The use of commercial expert system shell such as G2® from Gensym Corporation for reasoning, encoding, and storage of process know-how and laws of science (physical, chemical etc) to generate logical conclusion when combined with the process connectivity matrix will be a preferred approach to hard-coding knowledge representation. This will enable the knowledge base to be updated from time to time as deemed necessary.

Database Implementation

A relational database developed using database management system such as MS Access® or Oracle® to serve as repository for data generated from plant connectivity information and results from data-driven analysis will be viable alternative to in-memory storage and processing plant elements. This will improve systems response time and data throughput. One way of implementing this approach is shown Figure 140.

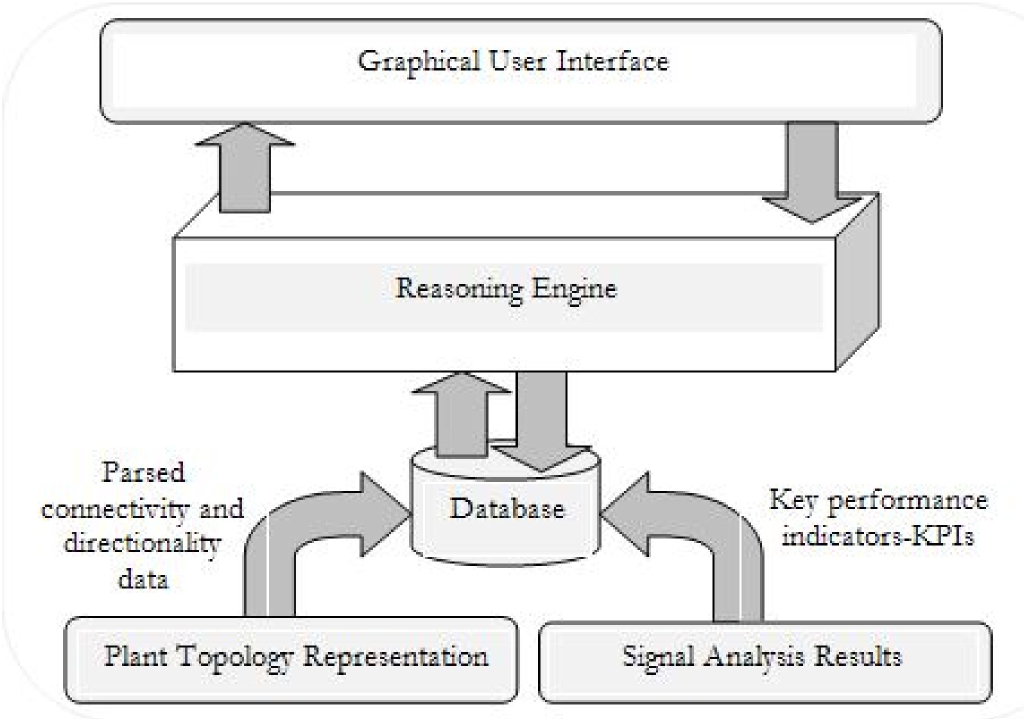


Figure 140: An alternative data-based implementation of Process Connectivity Analyser tool

7.4.2 Scale Up and Commercialization

Process Connectivity Analyser is designed and implemented with the possibility of integration into an existing commercial data-driven software tools in mind. The graphical user interface was developed using the windows presentation foundation (WPF) technology that easily integrates with other Windows based applications and highly modular such that the implementation code called the *engine* can be easily separated from the user interface elements.

The navigation-based graphical user interface presents a familiar Windows browser interface and can be easily adapted to a browser interface using XAML browser application (XBAP). XAML stands for extensible application mark-up language. Conversion of WPF to XBAP is straightforward requiring minimal developmental effort and allows the connectivity application to be accessed remotely from client software application such as Microsoft Internet Explorer or Firefox.

7.4.3 Augmented Intelligent P&IDs

Exports from process P&IDs as used in this project are static boolean representations that gives a yes or no (one or zero) answer to the existence or otherwise of connection between two pieces of plant items. This can be modified to incorporate other information for an improved functionality. This information could be numerical values derived from process measurements statistics such that the connectivity matrix entries are indicative of the information being conveyed as opposed to only boolean zeros and ones. For example, estimates of time delay from two instrument readings, if exists, can be used to estimate relative physical distance/separation between the measurement points.

Collaboration between intelligent CAD tool vendors and the academic could produce information rich P&IDs which can be used for enhanced process analysis. Findings from academic research could be fed to CAD vendors to incorporate addition information about plant entities and directional connections among plant items.

The discrete nature of the interconnections among plant entities represented by binary zeros and ones in the connectivity matrix entries can benefit from mixed integer non-linear optimization techniques (Kocis and Grossmann, 1989; Reyes-Labarta, *et al*., 2011; Turkay and

Grossmann, 1996; Viswanathan and Grossmann, 1990) when combined with continuous process variables.

Similarly, the information content of the connectivity matrix can be augmented if its elements are generated according to a certain rule such as addition of weight (Vianna and McGreavy, 1995) or dimension of connection to a node. The addition of extra information to the arcs will find application in future research work, for example by using the importance of the equipment or components attached to an arc to give a weight to such arc.

Another useful implementation will be the ability to manipulate XML description of process schematics algorithmically and export the manipulated XML file back into CAD tool to display process schematic graphically using the manipulated XML to reflect the changes in the XML. This can be used, for instance, for colour coding of various disturbance propagation paths and will require backward compatibility from intelligent CAD tools.

7.4.4 Sparse Matrix and Matrix Transforms

A vast number of entries in connectivity matrices generated from process schematics are zeros. These connectivity matrices can therefore benefit from sparse matrix algorithms that substantially reduce the storage requirements and search speed. A sparse matrix is populated mainly with zeros, the property that is utilized by a number of special techniques which stores and operate on non-zero entries (Mah, 1983). The basic idea is to devise an algorithm which minimizes the matrix fillings (Mah, 1974; Mah, 1983). The work reported by Jiang *et al*.,(2009) for example, might find application but developments might be needed to deal with scalability issues. For instance, in reachability matrix analysis, an N by N matrix will have to be raised to the power N, so there will be research issues related to calculation and compression methods for sparse matrices to make the graph theory approach feasible for practical large-scale applications.

7.5 Conclusions

This thesis has presented innovative and new ways to capture and manipulate information from large scale process schematic such as a refinery in order to give an automated means of diagnosing plant-wide performance problems and perform other process analysis.

The scale of the industrial case study described in Section 5.5.4 with part of its connectivity matrix shown in Figure 115 coupled with P&IDs merging capability proved the scalability and large-scale industrial relevance of the research concepts.

The practical outcome of the PhD work is an improved plant-wide disturbance diagnostics using process data, information from process intelligent P&ID and process understanding.

The procedures and technology for representing process plant electronically for the purpose extracting plant items and connectivity information in an industrial setting and scale have been demonstrated. A survey of commercial tools for automated electronic capture of process P&IDs and export, and tools for data-driven analysis was presented.

The work reported in this thesis has been engineered into a software tool that can innovatively access process intelligent P&ID in XML format, merge multiple P&IDs if required, and extract and combine engineering data with traditional process measurements and process understanding for resolving operational issues.

The industrial requirements for a large scale integrated plant-wide approach to process diagnostics utilizing all available information sources: measurement data, process topology information and process understanding have been addressed in this thesis.

Due to the confidentiality nature of the research, the technical reports generated from this research have not been made available in the public domain. In addition to technical reports produced, meetings were held with the sponsor's representatives on a regular monthly basis via teleconference and presentation on work done given quarterly.

8 References

ABB (2002) Expertly controlled, http://library.abb.com/global/scot/scot244.nsf/VerityDisplay/5A333DAE15510935C1256D60001CCEF5/$File/World_Cement_Jan2002.pdf, *Accessed 20 June 2011*.

AgileManifesto.org (2011) Manifesto for Agile Software Development, *http://agilemanifesto.org/ Accessed 30 June 2011*.

Alabi, D.B. (2010) Automated Analysis of Control Degree of Freedom. *Department of Chemical Engineering and Chemical Technology*. MSc Thesis, Imperial College London, London, pp. 78.

AlGhazzawi, A. and Lennox, B. (2008) Monitoring a complex refining process using multivariate statistics, *Control Engineering Practice*, 16, 294-307.

Arulampalam, M.S., Maskell, S., Gordon, N., and Clapp, T. (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking, *IEEE transactions on signal processing*, 50, 174-188.

Bauer, M., Cox, J.W., Caveness, M.H., Downs, J.J. and Thornhill, N.F. (2007) Finding the direction of disturbance propagation in a chemical process using transfer entropy, *IEEE Transactions on Control Systems Technology*, 15, 12-21.

Bauer, M. and Thornhill, N.F. (2008) A practical method for identifying the propagation path of plant-wide disturbances, *Journal of Process Control*, 18, 707-719.

Bearman, R.A. and Milne, R.W. (1992) Expert systems: Opportunities in the minerals industry, *Minerals Engineering*, 5, 1307-1323.

Beez, S., Fay, A. and Thornhill, N. (2008) Automatic Generation of Bond Graph Models of Process Plants, *2008 IEEE International Conference on Emerging Technologies and Factory Automation, Proceedings*, 1294-1301.

Bouamama, B.O., Medjaher, K., Samantaray, A.K. and Staroswiecki, M. (2006) Supervision of an industrial steam generator. Part I: Bond graph modelling, *Control Eng. Practice*, 14, 71-83.

Bouamama, B.O., Thoma, J.U. and Cassar, J.P. (1997) Bond graph modelisation of steam condensers, *Smc '97 Conference Proceedings - 1997 IEEE International Conference on Systems, Man, and Cybernetics, Vols 1-5*, 2490-2494.

Buede, D.M. (2009) *The engineering design of systems: models and methods*. John Wiley & Sons, Inc. NJ USA.

Chiang, L.H. and Braatz, R.D. (2003) Process monitoring using causal map and multivariate statistics: fault detection and identification, *Chemometrics and Intelligent Laboratory Systems*, 65, 159-178.

Choudhury, A.A.S., Shah, S.L. and Thornhill, N.F. (2008), *Diagnosis of process nonlinearities and valve stiction*, Springer-Verlag Berlin Heidelberg.

Christofides, P.D., Davis, J.F., El-Farra, N.H., Clark, D., Harris, K.R.D. and Gipson, J.N. (2007) Smart plant operations: Vision, progress and challenges, *AICHE Journal*, 53, 2734-2741.

Cochran, E., L. , Miller, C. and Bullemer, P. (2011) Abnormal situation management in petrochemical plants: can a pilot's associate crack crude?, *ASM Consortium*, http://www.asmconsortium.net/Documents/NAECONTC4.pdf. *Accessed 5 April 2011.*

Daratech (2004) Best Practices, *DBP-030401*: Implementing Intelligent P&IDs at Rayong Olefins and Lyondell Chemical Company. Cambridge, MA. http://www.daratech.com/research/bestpractices/intelligent_pid.pdf, *Accessed 4 April 2011.*

De Kleer, J. and Brown, J.S. (1984) A qualitative physics based on confluence, *Artificial Intelligence* 24: 7-83.

Desborough, L. and Miller, R. (2002) Increasing customer value of industrial control performance monitoring-Honeywell's experience, *AIChE Symposium series No 326, 98, 153-186.*

Di Geronimo Gil, G.J. (2010) Introducing First Principles Model Information in Plant Topology. *Department of Chemical Engineering and Chemical Technology*. MSc Thesis, Imperial College London, London, pp. 176.

Downs, J.J. and Vogel, E.F. (1993) A plant-wide industrial-process control problem, *Comput. Chem. Eng.*, 17, 245-255.

Edwards, V.A. and Whitaker, R.J. (2007) Fault management: A functional view of root-cause analysis and correlation. http://www.tavve.com/EW_White_Paper.pdf. Tavve software company, Morrisville, NC 27560. *Accessed 12 June 2010*.

ENR, M.H. (2005) http://enr.construction.com/aboutus/contact/FAQ.asp. *Accessed 30 June 2011.*

Epple, U., Fedai, M., Drath, R., Fay, A.(2002) Eine neutrale Beschreibung für die lebenszyklusbegleitende Spezifikation und Implementierung verfahrenstechnischer Anlagen auf der Basis von XML, *In: VDI-Berichte, No. 1684, VDI-Verlag, Düsseldorf, Germany, S. 133-143.*

Farrell, B.F. and Ioannou, P.J. (2001) State estimation using a reduced-order Kalman filter, *J. Atmos. Sci.*, 58, 3666-3680.

Fedai, M. and Drath, R. (2005) CAEX-A neutral data exchange format for engineering data, *ATP International Automation Technology 01/2005, 3, 43-51.*

Fedai, M. and Drath, R. (2005) CAEX-A neutral data exchange format for engineering data, *International Automation Technology* 01/2005, 3, 43-51.

Forbus, K.D. (1993) Qualitative process theory - 12 years after, *Artificial Intelligence*, 59, 115-123.

Frank, P.M. (1994) Online fault-detection in uncertain nonlinear-systems using diagnostic observers - a survey, *Int. J. Syst. Sci.*, 25, 2129-2154.

Frank, P.M., Garcia, E.A. and Koppen-Seliger, B. (2000) Modelling for fault detection and isolation versus modelling for control, *Mathematics and Computers in Simulation*, 53, 259-271.

Gao, D., Wu, C.G., Zhang, B.K. and Ma, X. (2010) Signed Directed Graph and Qualitative Trend Analysis Based Fault Diagnosis in Chemical Industry, *Chin. J. Chem. Eng.*, 18, 265-276.

Gartner (2008) http://www.gartner.com/technology/home.jsp. *Accessed 15 March 2011.*

Girardot, M. and Sundaresan, N. (2011) Millau: an encoding format for efficient representation and exchange of XMLover the Web, http://www9.org/w9cdrom/154/154.html*, Accessed 10 August, 2011*.

Gross, J.L. and Yellen, J. (2004) Handbook of graph theory. *Discrete Mathematics and Its Applications*. CRC Press LLC, Boca Raton, Florida.

Gunther, J.C., Conner, J.S. and Seborg, D.E. (2007) Fault detection and diagnosis in an industrial fed-batch cell culture process, *Biotechnology Progress*, 23, 851-857.

Harris, T.J., Seppala, C.T. and Desborough, L.D. (1999) A review of performance monitoring and assessment techniques for univariate and multivariate control systems, *Journal of Process Control*, 9, 1-17.

Horch, A. and Isaksson, A.J. (1999) A modified index for control performance assessment, *J. Process Control*, 9, 475-483.

Iri, M., Aoki, K.,Oshima, E. and Matsuyama, H. (1979) An algorithm for diagnosis of system failures in the chemical process, *Computers & Chemical Engineering* Volume: 3 Issue: 1-4 Pages: 489-493 DOI: 10.1016/0098-1354(79)80079-4

Iyun, O.E. (2005) An exploration of qualitative approaches to modelling and simulation of physical processes, MSc dissertation (p.79), University of Wales, Aberystwyth, United kingdom.

Jelali, M. (2006) An overview of control performance assessment technology and industrial applications, *Control Engineering Practice*, 14, 441-466.

Jiang, H.L., Patwardhan, R. and Shah, S.L. (2009) Root-cause diagnosis of plant-wide oscillations using the concept of adjacency matrix, *J. Process Control*, 19, 1347-1354.

Julier, S., Uhlmann, J. and Durrant-Whyte, H.F. (2000) A new method for the nonlinear transformation of means and covariances in filters and estimators, *IEEE Trans. Autom. Control*, 45, 477-482.

Kankar, P.K., Sharma, S.C. and Harsha, S.P. (2011) Fault diagnosis of ball bearings using machine learning methods, *Expert Syst. Appl.*, 38, 1876-1886.

Katipamula, S. and Brambley, M.R. (2005) Methods for fault detection, diagnostics, and prognostics for building systems - A review, part I, *Hvac&R Research*, 11, 3-25.

Kinney, T. (2005) Interaction Detection Using Oscillation Analysis, *ISA Expo 2005.*

Kocis, G.R. and Grossmann, I.E. (1989) A modeling and decomposition strategy for the minlp optimization of process flowsheets, *Comput. Chem. Eng.*, 13, 797-819.

Kokawa, M., Miyazaki, S. and Shingai, S. (1983) Fault location using digraph and inverse direction search with application, *Automatica*, 19, 729-735.

Koning, H. (2007) Harmonizing standardization on Data Integration in the life cycle of the Process and Power plants, http://www.wib.nl/files/miniseminar2008_item_3.pdf: *Accessed 20 June 2011*.

Kosebalaban, F. and Cinar, A. (2001) Integration of multivariate SPM and FDD by parity space technique for a food pasteurization process, *Computers & Chemical Engineering*, 25, 473-491.

Kramer, M.A. and Palowitch, B.L. (1987) A rule-based approach to fault-diagnosis using the signed directed graph, *AICHE Journal*, 33, 1067-1078.

Kuipers, B. (1986) Qualitative simulation, *Artificial Intelligence*, 29, 289-338.

Laud, A.R. (2011)
XMpLant and ISO 15926 presentation http://www.noumenon.co.uk/services.html. *Accessed 10 February 2011*.

Lee, G. and Yoon, E.S. (2003) Multiple-fault diagnosis using dynamic PLS built on qualitative relations, *European Symposium on Computer Aided Process Engineering - 13*, 14, 443-448.

Lee, G.B., Song, S.O. and Yoon, E.S. (2003) Multiple-fault diagnosis based on system decomposition and dynamic PLS, *Industrial & Engineering Chemistry Research*, 42, 6145-6154.

Lee, M.H. (1999) On models, modeling and the distinctive nature of model-based-reasoning, *AI Communications, Vol 12, Issue 3, pp. 127-133*.

Li, H.R., Braun, J.E. and Ashrae (2007) Economic evaluation of benefits associated with automated fault detection and diagnosis in rooftop air conditioners. In, *ASHRAE Transactions 2007 Vol 113, Pt 2*. Amer Soc Heating, Refrigerating and Air-Conditioning Engs, Atlanta, pp. 200-210.

Luyben, W.L., Tyreus, B.D. and Luyben, M.L. (1999), *Plant-wide process control*, New York: McGraw-Hill.

Lyman, P. R., and Georgakis, C. (1995) Plant-wide control of the Tennessee Eastman problem, *Computers & Chemical Engineering* vol. 19. no.3, pp. 321-331.

Mah, R.S.H. (1974) Constructive algorithm for computing reachability matrix, *AICHE J.*, 20, 1227-1228.

Mah, R.S.H. (1983) Application of graph-theory to process design and analysis, *Comput. Chem. Eng.*, 7, 239-257.

Maurya, M.R., Rengaswamy, R. and Venkatasubramanian, V. (2003) A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis, *Ind. Eng. Chem. Res.*, 42, 4789-4810.

Medvidovic, N., Rosenblum, D.S., Redmiles, D.F. and Robbins, J.E. 2002 Modeling software architectures in the unified modeling language. *ACM transactions on software engineering and methodology* Volume: 11 Issue: 1 Pages: 2-57 DOI: 10.1145/504087.504088.

Microsoft (2011) http://www.microsoft.com/net/, *Accessed 30 June 2011.*

Microsoft, C. (2006) Introducing the Office 2007 Open XML File Formats, http://msdn.microsoft.com/en-us/library/aa338205%28v=office.12%29.aspx., *Accessed 15 June 2011*.

MindGenius (2008) http://www.adeptscience.de/mindgenius/whitepapers/CauseEffectAnalysis.pdf; *Accessed 23 February 2008.*

Mohindra, S. and Clark, P.A. (1993) A Distributed Fault-Diagnosis Method Based on Digraph Models - Steady-State Analysis, *Computers & Chemical Engineering*, 17, 193-209.

Nandi, S., Toliyat, H.A. and Li, X.D. (2005) Condition monitoring and fault diagnosis of electrical motors - A review, *IEEE Transactions on Energy Conversion*, 20, 719-729.

Nimmo, I. (1995) Abnormal situation management. *Process and Control Engineering*, 49, 5, 8

NIST (2004) http://www.nist.gov/director/planning/summary_strategic_planning_study_results.cfm. *Accessed 30 June 2011*.

Norvilas, A., Norvilas, A., Negiz, A., DeCicco, J. and Cinar, A. (2000) Intelligent process monitoring by interfacing knowledge-based systems and multivariate statistical monitoring, *Journal of Process Control*, 10, 341-350.

Oyeleye, O. O. (1990) Qualitative modelling of continuous chemical processes and applications to fault diagnosis, *PhD Dissertation, Massachusetts Institute of Technology (MIT*), Boston USA.

Patton, R.J. and Chen, J. (1994) Review of parity space approaches to fault-diagnosis for aerospace systems, *J. Guid. Control Dyn.*, 17, 278-285.

Paulonis, M.A. and Cox, J.W. (2003) A practical approach for large-scale controller performance assessment, diagnosis, and improvement, *J. Process Control*, 13, 155-168.

PCA (2011) https://www.posccaesar.org/wiki/PCA/About. *Accessed 10 July 2011*.

PCA (2011) Introduction to ISO 15926., https://www.posccaesar.org/wiki/ISO15926Primer#, *Accessed 15 July 2011.*

Perry, R., Supomo, A., Mular, M. and Neale, A. (2000) Monitoring control loop health at PT Freeport Indonesia. *Mineral and Metallurgical Processing*. Soc Min Engineers Aime, Littleton.

Pinotti, R., Zanin, A.C. and Moro, L.F.L. (2008) Advanced Control Monitoring in Petrobras' Refineries: Quantifying Economic Gains on a Real-Time Basis. In Braunschweig, B. and Joulia, X. (eds), *18th European Symposium on Computer Aided Process Engineering*. Elsevier Science Bv, Amsterdam, pp. 495-500.

Prakash, J., Patwardhan, S.C. and Shah, S.L. (2011) On the choice of importance distributions for unconstrained and constrained state estimation using particle filter, *J. Process Control*, 21, 3-16.

Prett, D.M. and Morrari, M. (1986) Shell Process Control Workshop, *Butterworths, Stoneham, MA 01280.*

Puccia, C.J. and Levins, R. (1985) Qualitative modeling of complex systems, *Harvard University Press, pp. 1-11*.

Qin, S.J. (1998) Control performance monitoring - a review and assessment, *Computers & Chemical Engineering*, 23, 173-186.

Reyes-Labarta, J.A., Brunet R.,Caballero-Suarez, J. A.,Boer, D. and Jimenez, L. (2011) Integrating process simulation and MINLP methods for the optimal design of absorption cooling systems. In Pistikopoulos, E.N.G.M.C.K.A.C. (ed), *21st European Symposium on Computer Aided Process Engineering*. pp. 301-305.

Ricker, N.L. and Lee, J.H. (1995) Nonlinear modeling and state estimation for the Tennessee-Eastman challenge process, *Comput. Chem. Eng.*, 19, 983-1005.

Scherf, T. (2006) Automated HAZOP analysis based on a CAEX plant description, *Diplomarbeit Automatisierungstechnik thesis, Helmut-Schmidt Universitat, Hamburg.*

Schleipen, M., Drath, R. and Sauer, O.(2008) The system-independent data exchange format CAEX for supporting an automatic configuration of a production monitoring and control system *IEEE International Symposium on Industrial Electronics, vols 1-5 pages: 317-322*

Sharma, J. (1997) Qualitative reasoning: Modeling and simulation with incomplete knowledge - Kuipers,B. In, *Journal of Classification*. pp. 177-179.

Spirtes, P. (2010) Introduction to Causal Inference, *Journal of machine learning research*, 11, 1643-1662.

Struss, P., Provan, G., de Kleer, J.and Biswas, G. (2010) Special Issue on Model-Based Diagnostics, *IEEE Trans. Systems Man Cybernetics Part A-System and Humans*, 40, 870-873.

Thambirajah, J., Benabbas, L., Bauer, M., and Thornhill, N.F. (2009) Cause-and-effect analysis in chemical processes utilizing XML, plant connectivity and quantitative process history, *Computers & Chemical Engineering*, 33, 503-512.

Thornhill, N.F. (2005) Finding the source of nonlinearity in a process with plant-wide oscillation, *IEEE Transactions on Control Systems Technology*, 13, 434-443.

Thornhill, N.F., Cox, J.W. and Paulonis, M.A. (2003) Diagnosis of plant-wide oscillation through data-driven analysis and process understanding, *Control Engineering Practice*, 11, 1481-1490.

Thornhill, N.F. and Horch, A. (2007) Advances and new directions in plant-wide disturbance detection and diagnosis, *Control Engineering Practice*, 15, 1196-1206.

Thornhill, N.F., Huang, B. and Zhang, H. (2003) Detection of multiple oscillations in control loops, *Journal of Process Control*, 13, 91-100.

Thornhill, N.F., Oettinger, M. and Fedenczuk, P. (1998) Performance assessment and diagnosis of refinery control loops. In Pekny, J.F.B.G.E. (ed), *Third International Conference on Foundations of Computer-Aided Process Operations*. pp. 373-379.

Thornhill, N.F., Shah, S.L., Huang, B. and Vishnubhotla, A. (2002) Spectral principal component analysis of dynamic process data, *Control Engineering Practice*, 10, 833-846.

Turkay, M. and Grossmann, I.E. (1996) Logic-based MINLP algorithms for the optimal synthesis of process networks, *Comput. Chem. Eng.*, 20, 959-978.

Umeda, T., Kuriyama, T., Oshima, E., and Matsuyama, H. (1980) A graphical approach to cause and effect analysis of chemical-processing systems, *Chemical Engineering Science*, 35, 2379-2388.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., and Yin, K. (2003a) A review of process fault detection and diagnosis Part I: Quantitative model-based methods, *Comput. Chem. Eng.*, 27, 293-311.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., and Yin, K. (2003b) A review of process fault detection and diagnosis Part II: Quantitative model and search strategies, *Computers & Chemical Engineering*, 27, 313-326.

Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., and Yin, K. (2003c) A review of process fault detection and diagnosis Part III: Process history based methods, *Computers & Chemical Engineering*, 27, 327-346.

Vianna, R.F. and McGreavy, C. (1995) Qualitative modeling of chemical processes - a weighted digraph (wdg) approach, *Comput. Chem. Eng.*, 19, S375-S380.

Villez, K., Srinivasan, B., Rengaswamy, R., Narasimhan, S., Venkatasubramanian, V. (2011) Kalman-based strategies for Fault Detection and Identification (FDI): Extensions and critical evaluation for a buffer tank system, *Comput. Chem. Eng.*, 35, 806-816.

Viswanathan, J. and Grossmann, I.E. (1990) A combined penalty-function and outer-approximation method for minlp optimization, *Comput. Chem. Eng.*, 14, 769-782.

w3.org (2011) XML http://www.w3.org/XML, *Accessed 27 June 2011*.

Wakeman, S.J., Chung, P. H. W., Rushton, A. G., Lees, F. P., Larkin, F. D. and McCoy, S. A. (1997) Computer aided hazard identification: Fault propagation and fault-consequence scenario filtering. In, *Hazards XIII Process Safety - the Future*. Inst Chemical Engineers, Rugby, pp. 305-316.

Walker, V. (2009) Designing a process flowsheet, *AIChE, Chemical Engineering Progress p.15*.

Warnars, H.L.H.S. 2011 Object-oriented modelling with unified modelling language 2.0 for simple software application based on agile methodology. Behaviour & Information Technology Volume: 30 Issue: 3 Pages: 293-307 DOI: 10.1080/01449290903186231

Yim, S.Y., Ananthakumar, H. G., Benabbas, L., Horch, A., Drathb, R. and Thornhill, N. F. (2006) Using process topology in plant-wide control loop performance assessment, *Computers & Chemical Engineering*, 31, 86-99.

Yu, W., Wilson, D.I. and Young, B.R. (2010) Control performance assessment for nonlinear systems, *Journal of Process Control*, 20, 1235-1242.