

Plant-wide fault and disturbance screening using combined network centrality and information-theoretic causality measure analysis

by

Simon Jacobus Streicher

A dissertation submitted in partial fulfillment
of the requirements for the degree

Master of Engineering (Control Engineering)

in the

Department of Chemical Engineering
Faculty of Engineering, the Built Environment and Information
Technology

University of Pretoria
Pretoria

September 2018

SYNOPSIS

Finding the source of a disturbance in complex systems such as industrial chemical processing plants can be a difficult task and require a significant amount of engineering hours. In many cases, a systematic elimination procedure is considered to be the only feasible approach but can cause significant process upsets. Practitioners desire robust alternative approaches.

This study evaluates methods for ranking process elements according to the magnitude of their influence in a complex system. The use of data-driven causality estimation techniques to infer an information transfer network among process elements is studied. Graph centrality measures are then applied to rank the process elements according to their overall effect.

A software implementation of the proposed methods forms part of this work.

KEYWORDS: prioritisation, maintenance, fault detection, connectivity, causal mapping

ACKNOWLEDGEMENTS

I would like to acknowledge the guidance of my supervisor, Carl Sandrock, who not only pointed me to possible methods and tools for achieving many of the tasks involved but also encouraged me to streamline my approach and follow best practices.

I am thankful for the encouragement and support of my extended family, parents and sister as well as the friends I made while writing this dissertation.

Kate, thank you for your support and enduring my constant preoccupation with this work for too many years. You inspire me.

ACRONYMS

- AIS** Active Information Storage
- APC** Advanced Process Control
- APM** Advanced Process Monitoring
- CAEX** Computer Aided Engineering Exchange
- CCA** Canonical Correlation Analysis
- CCM** Cross Convergent Mapping
- CLEB** Control Loop Economic Benefit
- CMI** Conditional Mutual Information
- CPI** Controller Performance Index
- CV** Controlled Variable
- DAG** Directed Asyclic Graph
- DTE** Directed Transfer Entropy
- DV** Disturbance Variable
- EKF** Extended Kalman Filter
- FDA** Fischer Discriminant Analysis
- FDD** Fault Detection and Diagnosis
- FFT** Fast Fourier Transform

FOPDT First Order Plus Dead Time

GC Ganger Causality

HPC High Performance Computing

iAAFT iterative Amplitude Adjusted Fourier Transform

IAE Integrated Absolute Error

ISE Integrated Square of Error

ICA Independent Component Analysis

ITD Interaction in Time Domain

ITN Information Transfer Network

JIDT Java Information Dynamics Toolkit

KPI Key Performance Indicator

k-NN *k* - Nearest Neighbour

KSG Kraskov-Stügbauer-Grassberger

MPC Model Predictive Control

MTR Multiple Time Region

MV Manipulated Variable

PCA Principal Component Analysis

PDF Probability Density Function

PFD Process Flow Diagram

PID Proportional Integral Derivative

P&ID Pipeline and Instrumentation Diagram

PLS Partial Least Squares

PSCI Power Spectral Correlation Index

PTE Partial Transfer Entropy

PV Process Variable

QTA Qualitative Trend Analysis

RBC Reconstruction Based Components

RGA Relative Gain Array

RPM Reset Probability Matrix

SDG Signed Directed Graph

SPC Statistical Process Control

SVD Singular Value Decomposition

SVM Support Vector Machines

TE Transfer Entropy

TPM Transition Probability Matrix

XML Extensible Markup Language

SYMBOLS

Greek symbols

α Katz attenuation factor

δ Optimal estimation delay

θ Transport delay

λ Eigenvalue

τ Time constant

ϕ Cross-correlation

ψ Transfer entropy

ω Directionality index

Latin symbols

A Adjacency matrix (binary)

e Natural number

G Transition probability matrix

H Entropy

I Mutual information

m Transition and reset probability mixing weight

N Number of samples

M Fully connected weight matrix

Q Reset probability matrix

R Reachability matrix

T Transfer entropy

W Ranking weight matrix

x Eigenvector

CONTENTS

Synopsis	i
Acknowledgements	ii
Acronyms	iii
Symbols	vi
1 Introduction	1
1.1 Fault detection in industry	2
1.2 Problem statement	3
1.3 Method	4
1.3.1 Connectivity	4
1.3.2 Node ranking	5
1.4 Contribution	5
1.5 Implementation details	6
1.6 Structure of this dissertation	6
I Theoretical development	7
2 Research framework	8
2.1 Fault management	8
2.2 Information based fault detection	10
2.3 Information theory and decomposition	11

2.4	Information manipulation in chemical processes	14
2.5	Probability density functions	15
2.6	Caveats	16
2.6.1	Noise	16
2.6.2	Scaling	16
2.6.3	Sampling frequency	17
2.6.4	Differentiating between disturbances and faults	17
2.6.5	Differentiating node importance	18
2.7	Motivation of selected approach	18
3	Structural modelling of processes using graphs	19
3.1	Applications in process engineering	20
3.2	Developing digraph models of processes	22
3.3	Statistical interaction measures	23
4	Quantifying causal connections from time series data	26
4.1	The notion of causality	26
4.2	Methods for causality estimation	28
4.2.1	Distance correlation	28
4.2.2	Directed partial correlation	28
4.2.3	Convergent cross mapping	28
4.2.4	Transfer entropy	29
4.3	Estimating probability density distributions	30
4.4	Available software packages	32
4.5	Choice of methods	33
4.6	Formal definitions of measures used	34
4.6.1	Transfer entropy	34
4.6.2	Cross-correlation	35
4.7	Significance testing	36
4.7.1	Transfer entropy	36
4.7.2	Cross-correlation	38
4.8	Scaling	40

5	Ranking process elements	42
5.1	Overview of prior art	42
5.2	Matrix representation of a directed graph	45
5.3	Centrality measures on a directed graph	45
5.3.1	Degree centrality	46
5.3.2	Closeness centrality	46
5.3.3	Betweenness centrality	46
5.3.4	Eigenvector centrality	47
5.3.5	Katz centrality	48
5.4	Markov chain interpretation	48
5.5	Incorporating meta-data in the ranking problem	49
5.6	Reachability analysis using an adjacency matrix	50
5.7	Direction of analysis	51
5.8	Normalisation	52
5.9	Implementation of ranking algorithms	53
5.10	Network reduction	53
II	Exploration	55
6	Influence of time series properties	56
6.1	Time constant	58
6.1.1	Kernel vs. KSG estimation	60
6.1.2	Effect of auto-embedding	61
6.1.3	Effect of relative time captured by sample	61
6.2	Noise generator sampling interval	62
6.3	Noise generator variance	62
6.4	Sample size	65
6.5	Sampling rate	65

7	Ranking methods and parameters	71
7.1	Incorporating indirect connections	71
7.2	Dummy nodes	72
7.3	Sample rankings	73
7.3.1	Eigenvector centrality	73
7.3.2	Biased eigenvector centrality	74
7.3.3	Katz centrality	75
7.3.4	Dummy nodes	75
8	Experimental analyses	78
8.1	Experimental design	78
8.2	Simulated level and composition control problem	79
8.3	Disturbances	80
8.4	Network impact rankings over time	83
8.5	Connectivity graphs	89
III	Implementation	93
9	Implementation and procedure	94
9.1	Software installation	94
9.2	Input files	94
9.3	Parameter selection	94
9.3.1	Sample size	95
9.3.2	Sampling rate	96
9.3.3	Delay optimisation range	97
9.3.4	Scaling approach	98
9.4	Computation time	98
9.5	Recommended analysis procedures	99
9.5.1	Initial screening procedure	99
9.5.2	Detailed root source detection for hypothesis formulation	102
9.5.3	Hypothesis testing	104

10 Tennessee Eastman case study	105
10.1 Problem overview	105
10.2 Choice of analysis parameters	111
10.3 Reactor cooling water temperature disturbance	111
10.4 Feed composition disturbance	115
10.5 Multiple time region analysis	118
11 Conclusion	120
11.1 Overview of the contributions of this work	120
11.1.1 Software package development	120
11.1.2 Motivation for applying information theory to FDD	120
11.1.3 Method and parameter selection guidelines	121
11.1.4 Unique elements	121
11.2 General conclusion	121
12 Recommended future research	123
12.1 Causality estimation and weight determination	123
12.1.1 Transfer entropy history embedding dimensions	123
12.1.2 Directionality and multiple variable effects	124
12.1.3 Power spectrum based initial screening and importance	125
12.1.4 Other measures	125
12.2 Network centrality measures	125
12.2.1 Application of additional centrality measures	126
12.2.2 Clustering elements as disturbance and fault sources	126
12.2.3 Incorporating meta-data in importance score calculations	127
12.3 Background and parallel computing	127
12.4 Multiple user collaborative web interface	128
12.5 Visualisation	128

A	Example process description	129
A.1	Process design parameters	129
A.2	Control limits	130
A.3	Controller tuning	130
A.4	Simulation and disturbance details	131
B	Software installation	132
B.1	Prerequisites	132
B.2	Installation and testing	133
C	Configuration, data and result formats	134
C.1	Configuration files	134
C.1.1	Path configuration file	134
C.1.2	Weight calculation configuration file	135
C.2	Time series data	144
C.3	Optional: Adjacency matrix	144
C.4	Output organisation	145
D	Detailed Tennessee Eastman ranking results	147

LIST OF FIGURES

5.1	Example directed graph with associated adjacency matrix	46
6.1	Interaction weights over a range of time delays for different process time constants	59
6.2	Interaction weights over a range of time delays for different process time constants that far exceed the time span covered by analysis	63
6.3	Interaction weights over a range of time delays for different noise sampling rates	64
6.4	Interaction weights over a range of time delays for different sample sizes .	66
6.5	Interaction weights over a range of time delays for different sampling rates	67
6.6	Interaction weights over a range of time constants for different sampling rates	68
6.7	Interaction weights over a range of time constants for different sampling rates close to the noise generation interval	70
8.1	Level and composition problem process flow diagram	80
8.2	High density time series plot of disturbance [standardised]	81
8.3	High density time series plot of disturbance [Skogestad scaled]	82
8.4	Multiple time region node rankings [simple; no significance testing] . . .	85
8.5	Multiple time region node rankings [directional; no significance testing] .	86
8.6	Multiple time region node rankings [simple; significance tested]	87
8.7	Multiple time region node rankings [directional; significance tested] . . .	88

8.8	Reduced edge ITNs for x_{sp} disturbance [simple; significance tested]	90
8.9	Reduced edge ITNs for x_{sp} disturbance [directional; significance tested]	91
10.1	Tennessee Eastman process flow diagram	106
10.2	High density time series plot of disturbances [standardised]	109
10.3	High density time series plot of disturbances [Skogestad scaled]	110
10.4	Transfer entropy information flow network for cooling water inlet temperature disturbance in Tennessee Eastman process [kernel estimator; directional weights; Skogestad scaled; bidirectional delays; significance tested]	112
10.5	Tennessee Eastman schematic with cooling water inlet temperature disturbance effects highlighted	113
10.6	Transfer entropy information flow network for cooling water inlet temperature disturbance in Tennessee Eastman process [KSG estimator; Skogestad scaled; bidirectional delays; significance tested]	114
10.7	Transfer entropy information flow network for stripper feed composition disturbance in Tennessee Eastman process [KSG estimator; Skogestad scaled; bidirectional delays; significance tested]	116
10.8	Multiple time region absolute ranking scores for periods covering both reactor CW inlet temperature and stripper feed composition disturbance in Tennessee Eastman process [kernel estimator; directional weights; Skogestad scaled; forward delays; significance tested]	119
A.1	Level and composition problem process flow diagram	129
D.1	Reactor cooling water disturbance results [cross correlation]	148
D.2	Reactor cooling water disturbance results [kernel estimator; standardised]	149
D.3	Reactor cooling water disturbance ranking results [kernel estimator; Skogestad scaled]	150
D.4	Reactor cooling water disturbance ranking results [KSG estimator; standardised]	151
D.5	Reactor cooling water disturbance ranking results [KSG estimator; Skogestad scaled]	152

LIST OF TABLES

2.1	Information theoretic measures	14
6.1	Base case parameters	57
7.1	Eigenvector centrality node importance distribution for chain	74
7.2	Eigenvector centrality node importance distribution for chain with bias	74
7.3	Katz centrality node importance distribution for chain with $\alpha = 3$	75
7.4	Dummy node demonstration gain matrix (no dummies)	76
7.5	Dummy node demonstration gain matrix (with dummies)	76
7.6	Eigenvector node importance distribution for chain with dummies	77
8.1	Weight analysis configuration options	79
8.2	Analysis configuration	79
8.3	Level and composition process variables	80
10.1	Tennessee Eastman problem relevant tags	107
10.2	Transfer entropy ranking results for CW inlet temperature disturbance [kernel estimator; directional weights; Skogestad scaled; forward delays; significance tested]	112
10.3	Cross correlation ranking results for CW inlet temperature disturbance [Skogestad scaled; significance tested]	115
10.4	Transfer entropy ranking results for feed composition disturbance [kernel estimator; directional weights; Skogestad scaled; forward delays; signifi- cance tested]	117

10.5 Cross correlation ranking results for feed composition disturbance [Skogestad scaled; significance tested]	118
A.1 Process variable control limits	130
A.2 Control loop tuning parameters	130

CHAPTER 1

INTRODUCTION

Control specialists working on large chemical processing plants are typically responsible for monitoring and optimising several hundred base layer control loops managing economic and safety variables such as flows, pressures and temperatures. The sheer number makes it nearly impossible to ensure that all control loops are operating satisfactorily at all times (Rahman and Choudhury, 2011). As Advanced Process Control (APC) technology and Model Predictive Control (MPC), in particular, matured and few obvious opportunities remain unexplored, a significant portion of the collective research effort in this field started moving towards the much broader scope of process monitoring.

The application of process monitoring techniques almost always involves an expert reviewing actionable advisories before execution of the recommended remedial action in the field. The robustness criteria for such operator-in-the-loop monitoring techniques are not as strict as for the case of automatic closed-loop process control, allowing significantly more advanced algorithms from the realms of machine learning and information theory to be implemented.

Various research groups suggested streamlined fault detection paradigms in recent years, significantly reducing the effort required to identify faulty process elements. However, it is not uncommon to find a backlog of already identified and known faults in need of attention, as the personnel is overwhelmed by the sheer volume of defects indicated in their areas of responsibility. Automated tools for calculating individual loop performance indicators exist, but analysing loops in isolation is unlikely to indicate loops that have the most significant impact on the plant's performance as a whole. Bauer (2005) identified the need for a triaging measure to prioritise control loop maintenance that can correctly report a regularly updated list of most pressing base layer control loops in need of expert

attention. Resolving the primary faults on processes can have a significant impact on product quality variability and may even make some proposed process equipment modifications unnecessary. However, without a clear idea of the most lucrative problem areas, control specialists' days are often spent "fighting today's fire", which is usually dictated by whatever happened to catch management's attention.

1.1 Fault detection in industry

The fields of process monitoring and Fault Detection and Diagnosis (FDD) are closely related to this project. Base layer control maintenance is an ongoing process. Reasons for control loops failing to meet performance objectives include:

- Controllers that were never tuned to meet necessary robustness and performance criteria, but merely to provide good enough performance to allow commissioning of the process for the first time.
- Poor management of change when process feed or conditions change, or an extensive modification is performed, such as the implementation of a heat integration network.
- Hardware failures and maintenance related issues, such as control valve hysteresis or sticking, or the clogging of heat exchangers, the deactivation of catalysts, and other discrete and gradual changes in process equipment behaviour.

Fault diagnosis is typically one of the most time-consuming activities that fall under a control engineer's list of responsibilities. Various disturbances and faults tend to be very hard to correct as it is laborious to trace the problem to its real source, resulting in many man hours lost chasing a faulty hypothesis brought about by false positives indicated by inadequate fault diagnosis methods. As it is regularly the case that multiple control loops' performances can suffer from a few core issues, an efficient method for identifying underlying causes is key. Experienced control engineers usually rely on an intuition that was built over many years to determine the most likely cause. A significant portion of this intuition is sourced from only knowing how the different control loops are interconnected, and thus able to interact with each other, propagating disturbances throughout the process. For example, it is quite common to find that the source of a quality specification being violated lies in feed upsets or poor control in recycle streams. Ranking control loops according to a measured degree of interaction is therefore considered to be an important yet challenging task (Rahman and Choudhury, 2011). The provision of connectivity and interaction metrics that can assist in troubleshooting procedures that previously relied on intuition acquired by experience is expected to be a significant aid in

maintaining contingency on plant operations. This is of particular interest in cases where the previous generation of operators is retiring in rapid succession.

Significant control loop performance deterioration related to a mechanical failure of some sort is usually quickly identified by panel operators who will subsequently place the loop in manual. However, panel operators are very likely to switch any loop that they feel uncomfortable with to manual control. A simple fault detection method based on the flagging of loops placed in manual mode by panel operators combined with the identification of loops that fail to maintain set points within specified margins may already identify more loops than can be handled by available personnel.

A good ranking tool will rank under-performing control loops such that a controller which is affecting the most important variables in the system is assigned higher priority than a control loop which may be under-performing to a greater extent but affecting less important variables. The indicated control loops should be as close to the cause of the fault as far as sensor placement allows, even if significantly amplified by another control loop.

Many fault detection tools require planned excitations of the process, a benchmark of normal fault-free operation and labelled data of known faults. These specifications are often undesirable or impractical in industry. An optimal method should require minimal initial configuration and maintenance effort to encourage the widespread adoption of the proposed solution by already overburdened control personnel. Ongoing operations should be automated and unsupervised.

A tool that provides accurate, regularly updated indicators of the process elements closely associated with the root causes of the most significant faults on a plant-wide scale would potentially have a drastic impact on the productivity of control teams. This work studies the usability of concepts from the fields of information and graph theory for incorporation into such a tool.

1.2 Problem statement

The problem addressed in this dissertation is the *prioritisation* of base layer control loops and other plant elements from a *maintenance* perspective by identifying factors that have the greatest impact on overall plant-wide profitability and safety. The method is *model-free* and *data-driven*.

The goal is to *identify, develop, implement* and *test* a method that is:

Accurate

The method must be correct about which loops are in the greatest need of mainte-

nance by directly indicating the most primary causal factors and filtering out effects from standard plant operation.

Efficient

Although detailed performance requirements will not be specified, as a rough guideline it is desired that several hundred process elements should be ranked in a 24 hour period by a server with a maximum rated power consumption of 1 kW.

Low maintenance

Both the initial implementation as well as the continued application of the method should be low in cost concerning quantity and skill level of man hours required. Experts are still expected to review the results before recommending maintenance steps.

1.3 Method

The method developed in this work focuses on analysing distributed effects of individual control loop performance or disturbances on a plant-wide scale. A two-step process is followed. First, the interaction between various elements in the process is determined using data-driven information-theoretic techniques. This connectivity information is used to create a directed graph (digraph) representation of the process, also referred to as an Information Transfer Network (ITN).

The second step calculates importance scores for all elements in the ITN according to their total distributed influence via application of graph centrality measures. To identify the control loops whose maintenance is likely to be the most lucrative investment of engineering and instrumentation specialist man-hours in a plant-wide context, quality information on individual loop performance and the economic value of the affected streams will also be required. Integration of this information is outside the scope of this work but briefly discussed in Future Research (Chapter 12).

1.3.1 Connectivity

Determination of connectivity between process variables is possible in at least two different ways: 1) measuring causal linkages using data-driven methods and 2) defining connections by making use of known process topology (referring to physical and logical connectivity) information contained in Pipeline and Instrumentation Diagram (P&ID)s which are normally kept well organised and reasonably up to date.

Data-driven methods have the advantage of being less labour-intensive. Disadvantages include the indiscriminate indication of higher-order degree-of-separation connections and

the generation of spurious results. Addressing these concerns are active topics in current literature. As it is desirable to have a digraph consisting of valid, single degree-of-separation connections to track the propagation path of interactions inside the plant accurately, the creation of a digraph is commonly approached by a hybrid method whereby the data-driven results are supervised and filtered according to knowledge of the plant's P&IDs.

As this work primarily focuses on the ranking of nodes with individual path tracing being only a secondary goal, the scope is limited to data-driven methods only, and transfer entropy in particular, for reasons given in Chapter 4, which also discusses the details of transfer entropy used in this dissertation, including optimisation with respect to time delays and the correct interpretation of values obtained.

1.3.2 Node ranking

A modified form of the Google PageRank algorithm (Bryan and Leise, 2006), which is essentially a graph eigenvector centrality measure, is used to rank control loops based on their connectivity, interaction and importance scores (weights). Farenzena and Trierweiler (2009) first applied this concept to the ranking of control loops and dubbed the algorithm LoopRank in this context. This work builds on the work of Wilken (2012) who also investigated the application of graph ranking to control monitoring and fault prioritisation applications.

1.4 Contribution

The contribution of this dissertation and related work is as follows:

- An overview of the most significant developments in various fields with the potential for addressing the problem of ranking process elements and disturbances according to their influence on overall plant performance is provided.
- A full development and analysis of the use of selected methods for the purpose of FDD is presented.
- Guidelines and procedures for methods as well as typical parameters to use in practical implementation is developed.
- A software package implementing the recommended methods is made available.
- Results for various tests and case studies generated using the developed software is discussed.

1.5 Implementation details

The core implementation of the method is done in the Python programming language (version 3.6) but makes use of some external libraries, most notably the Java Information Dynamics Toolkit (JIDT) developed by Lizier et al. (2008). Other open source tools such as Cytoscape (Shannon et al., 2003) are used to visualise results. The software runs on Linux as well as Microsoft Windows operating systems, and a Docker container is available for easy deployment.

Supporting files are available digitally and should be accessible together with a distribution of this dissertation. A git repository containing source code produced during this study is accessible at: <https://github.com/SimonStreicher/FaultMap.git>

1.6 Structure of this dissertation

The remainder of this document is structured in three parts:

Part I (Theoretical development) reviews prior art related to the fields of information theory (Chapter 2), functional modelling and plant connectivity (Chapter 3), causal interaction inference (Chapter 4) and graph ranking algorithms (Chapter 5). The selection of some these methods for further investigation is motivated.

Part II (Exploration) discusses the effect of signal properties on the behaviour of different causal interaction metrics (Chapter 6), the relationship between network structure, edge weights and relative ranking scores (Chapter 7) and demonstrates typical results on a simple simulated example (Chapter 8).

Finally, Part III (Implementation) provides a recommended procedure for approaching a fault diagnosis problem using the discussed techniques (Chapter 9), a detailed case study on the Tennessee Eastman problem (Chapter 10), a summary of the main contributions and observations that result from this work (Chapter 11) and an overview of aspects considered worthy of further investigation (Chapter 12).

Part I

Theoretical development

CHAPTER 2

RESEARCH FRAMEWORK

This chapter presents the basic premise from information theory upon which the function and desired output of a purely data-driven complex system analysis and monitoring tool are based and explores the theoretical upper bound of such a tool's performance and utility. Also, this chapter contains a gentle introduction to information theory and decomposition and briefly discusses some caveats involved to adjust for the unique nature of chemical processing data. The selection of specific measures and techniques studied in this dissertation is motivated. Chapters 4 and 6 that follow provide further details of the selected methods.

2.1 Fault management

Fault management is an important aspect of any industrial process' operation, allowing for effective management of abnormal events and striving towards optimal performance. It is, therefore, no surprise that this is a field of considerable interest to many research groups worldwide, with significant advances made at a rapid pace. The Eastman Chemical Company successfully implemented a large-scale controller performance assessment system with the aim of ranking process control issues to prioritise maintenance (Paulonis and Cox, 2003). The ever-expanding fields of data science, mining, machine learning and artificial intelligence combined with the readily available computational infrastructure and resources required play a significant role. A global trend of increased collaboration among researchers in unrelated industries led to the identification of common goals and desires, resulting in the rapid adoption and development of new techniques.

The field of fault management in process engineering consists of two sub-disciplines referred to as fault detection and diagnosis respectively. Detection is concerned with indicating whether and where control upsets occur and fall in the broader category of process monitoring. On the other hand, fault diagnosis relates to the identification of optimal resolutions of deviations from control limits. As such it can be considered to be an aspect of general process troubleshooting.

The primary focus of this work, i.e. triaging control underperformance on a plant-wide scale, can be deemed to be concerned mostly with fault detection. However, since it incorporates methods typically proven to be useful in fault diagnosis, these aspects will also be discussed. The problem of ranking the most critical process elements according to their influence on a plant's performance as a whole is naturally closely related to the problem of finding causes of individual process faults.

It is helpful to distinguish between faults that originate outside and inside control loops. Individual loop Key Performance Indicator (KPI)s are widely accessible, making it simpler to triage faults originating inside control systems. Also, failures arising inside control loops require significantly less capital expenditure to correct on average as the required remedial action is often limited to parameters changes in software compared to the adoption and replacement of process equipment, for example. The method presented in this work aims to indicate the most significant process element and is not necessarily limited to control loops only, and will, therefore, deal with both of these categories to some degree.

Venkatasubramanian et al. (2003) published a thorough review series on process fault detection and diagnosis methods, while Thornhill and Horch (2006) presents an excellent discussion of plant-wide controller performance assessment. However, many of these methods require either a database of known faults or a benchmark of nominal, fault-free operation. These requirements are a significant barrier to entry for widespread industrial adoption use where it is often critical that advanced analysis methods be robust and easy to implement and maintain.

Two categories of fault diagnosis methods are (1) those that require a known model of the process, usually generated by extensive use of process knowledge obtained from experts, and (2) data-driven techniques. Data-driven methods are more practical as opposed to methods that rely on rigorous process models which rarely available in industry (Liu and Chen, 2014).

2.2 Information based fault detection

Consider a chemical process under automatic closed-loop control. The time-series signals captured from periodic sampling of measurement devices from such a system can be expected to show behaviour that is:

- continuous,
- dynamic,
- multivariate,
- highly interactive,
- complex,
- non-linear.

Systems with the above properties are relatively challenging to analyse robustly compared to discrete, stationary, univariate and linear systems, for which the theory is well developed.

Some of the signals sampled from a chemical process indicate information of high value; the most critical states would typically be of significant economic or safety concern. Detecting deviations from optimal operation using Advanced Process Monitoring (APM) and FDD methods is an active (and potentially lucrative) industrial and academic research interest.

An APM system will monitor all process states in near real-time, drawing attention to the most critical behaviour occurring at any point in time. Advanced systems might provide insight into the causal flow of events that results in said behaviour and offer actionable advisories such that operating personnel can take optimal corrective remedial action.

One key feature distinguishes the problem of chemical process monitoring from most other systems of interest that exhibit similar behaviour, such as the stock market: chemical processes are inherently deterministic. Every single measurement reflected by every probe and each floating point value stored in a process historian is ultimately the result of a deterministic physical system that follows a set of heavily convoluted but individually simple laws.

The proposed framework of fault detection presented relies on the premise that the amount of uncertainty introduced in a system that impacts significant variables indicates the severity of a fault or disturbance. The method is further motivated by the belief that a combination of information and graph theory may assist in tracing this uncertainty back to its source.

2.3 Information theory and decomposition

Consider the sampled time series signal from a single measurement of a process state, such the temperature on a distillation tray. The process state at each sampling instance is not entirely independent of the previous states, and as such, every new measurement is less informative than that of a white noise signal, for which it is not possible to infer a probable range for the value from previous values. For physical systems that obey a set of fundamental rules, the signal are expected to exhibit behaviour that is less than utterly random.

Shannon (1948) described the information content of a single signal regarding the predictability of its values, with predictable signals containing less information than more seemingly random ones. From another perspective, Shannon entropy quantifies the information content of a single signal in terms the minimum bits required to fully reconstruct a signal under optimal compression.

A signal's Probability Density Function (PDF) provides a basis for interpreting its uncertainty. Discrete signals with countably limited options, it is easy to compute the PDFs precisely. For continuous variables not modelled according to a specific statistical distribution with an analytical solution for entropy, the accuracy will depend on the suitability of the method used to estimate the PDF. The bandwidth and/or other parameters used by the estimator will influence the absolute value returned.

Most of the activity described by the sampled time series of chemical processes are deterministic consequences of intended interactions, together with some amount of novel information uniquely represented in this measurement. In this discussion, all sources of original information in a single measurement is considered to belong to one of the following categories:

Exogenous inputs Inputs from other states (nodes) in the network.

Disturbances External factors not included in one of the states observed for the system under investigation. Nodes exclusively influenced by factors external to the battery limits of the system will receive the bulk of their information content via external disturbances.

Measurement noise Imperfect measurement is expected to add some entropy to a variable's time series signal. For this reason, it might be a good idea to perform state estimation as a preliminary step before performing any further inference in an attempt to partition measurement noise from the actual state. If measurement noise occurs on a controlled variable such that it enters the control algorithm, it will become a real disturbance to the manipulated variable.

Faults A fault in the process that links two nodes can cause additional entropy. For example, cavitation over a valve will cause the flow signal to have more entropy than expected from a simple non-linear relationship between upstream pressure and valve position.

Faults and disturbances can introduce new information to any state in a process. To trace the entropy of observed variables uniquely to its sources, the algorithm must be careful not to count information twice.

Suppose it was possible to perform an “information balance” on a system in a similar fashion to an energy or mass balance, tracing the proportional information content contained in any time series signal used to monitor the state of a system back to all of its exogenous inputs, measurement noise, disturbances and faults. Observing how these dynamics change over time would be a potent tool for identifying any deviations caused by faults or abnormal disturbances.

However, unlike the rules of mass and energy balances, where the quantity of interest cannot be created or destroyed but merely transferred and transformed, information content plays by a different set of rules that do not allow for simple arithmetic operations to yield accurate results, making the account problem significantly more complicated.

Information can be:

- created,
- destroyed,
- transferred,
- stored,
- processed,
- modified through redundant and synergistic interactions.

Partitioning the entropy of a signal into a measure of relative contributions from other sources and interactions is known as information decomposition. Faes et al. (2016) describes a complete approach for dissecting the information of a target variable, using either variance or entropy as a basis. Their approach uniquely divides signal information into:

- new information produced by the system,
- information stored in the system,
- information transferred to it from other systems.

Information transferred is further partitioned into the specific contribution of measured sources to the target dynamics and amounts reflecting information modification. Modi-

fication refers to the balance of redundant and synergetic interactions between elements of the system.

Even a simplified model of a typical chemical plant will most likely involve hundreds to thousands of states. Rigorous information decomposition for a general non-linear process from sampled data only (no analytical models) becomes computationally intractable past a few states. On the other hand, Faes et al. (2017) provides an analytical solution to the information decomposition problem for multivariate Gaussian processes.

Various researchers published attempts at addressing the exploding complexity, including a greedy algorithm for calculating all the information transferred to a single variable known as complete transfer entropy (Lizier and Rubinov, 2012). As a second prize to full multivariate information decomposition, complete bivariate information transfer analysis combined with graph theory can provide a sufficient approximation to allow for causal reasoning (Runge et al., 2012).

Simplifying things further, partial bivariate information transfer metric combined with graph centrality measures can indicate the most important sources of original information, without necessarily providing an accurate representation of the unique and direct internal information pathways in the system. This work aims to develop and study the usefulness of a plant-wide fault triaging indicator and not provide causal reasoning support per se. This approach that is two steps away from the near-unattainable theoretical optimum is expected to provide an appropriate balance between rigour and computational complexity.

As such, the only information theoretic metric involved will be bivariate transfer entropy as presented in Section 4.2. An accurate estimation of transfer entropy indicates the amount of uncertainty a source signal transferred to a destination node's observed signal. Table 2.1 places some relevant information theoretic measures and the complexity of transfer entropy in context Cover and Thomas (2006); Faes et al. (2016).

The concept of entropy is defined for discrete signals with a countably finite amount of possible states, while the uncertainty of continuous signals is described by differential entropy and is estimated from PDFs Cover and Thomas (2006).

Table 2.1: Information theoretic measures

Entropy	A measure of uncertainty in a signal. Observations of uncertain states carry more information than samples from predictable signals.
Conditional entropy	The remaining entropy after accounting for uncertainty introduced due to the effect of other processes.
Mutual information	The entropy not unique to a specific signal but shared with other signals evaluated.
Conditional mutual information	Remaining entropy shared among a set of signals after accounting for the effect of other processes.
Transfer entropy	The entropy shared between a source and target signal after conditioning on the target signal's past; a measure of conditional mutual information in time.
Total transfer entropy	The entropy <i>uniquely</i> shared between a source and target signal measured by conditioning on every other signal in the system under investigation in addition to the target's past.

2.4 Information manipulation in chemical processes

As stated before, chemical processes can be considered deterministic, with the information transfer between nodes happening according to a some (linear or non-linear) transform. Such transforms cannot create original information.

If the differential entropy of states exciting a node is more than the incoming entropy, the additional information must have originated from noise or a process fault. However, transferring processes are likely to destroy particular parts of the incoming information content, and therefore not all the differential entropy of a source node connected via some process to a target node will have an equal impact. As an extreme example, the entropy of a time signal coming out of zero-order hold process is zero (it is entirely predictable), while a First Order Plus Dead Time (FOPDT) process will act as a low-pass filter, attenuating information contained in high frequencies.

Chemical processes typically contain numerous signals that are highly co-linear. This multiplicity of information is a concern during the selection of control and manipulated variables in the design of advanced model-based controllers as its presence can lead to apparent conflict in the objective function and make the controller sensitive to plant and model mismatch. Collinearity introduces the potential information double-counting in decomposition restricted to bivariate analysis.

Consider a process that merely sums all its input signals without delay. This process does not inherently destroy information, and if it received a single input signal, the entropy of its output would be equal to that of the input signal. In the case of multiple signals entering the process, remaining entropy of the output signal will depend on the shared information between the input signals. If the input signals were identical, output signal will have the same entropy as any of the individual input signals, since multiplying a signal by a constant gain does not impact its entropy. Mutual information is an entropy-based equivalent for the variance based concept of collinearity and only contributes once to the output. The output signal entropy of a summing node will be equal to the mutual information of the input signals plus the sum of unique entropies in the source signals.

2.5 Probability density functions

The PDF of a signal indicates the probability of a random single observation being a specific value. For example, finding values near unity on the absolute magnitude of a pure sine signal is less probable than values close to zero as the signal spends more of its time close to values near zero than away from it.

The accuracy of the method used to estimate the PDF of a signal limits the accuracy of a signal's entropy estimate. Exact PDF calculation of discrete signals with finite possible values is straightforward and merely involves counting the occurrence of each value and dividing by the total number of samples.

The entropy of continuous signals that follow certain analytic distributions, such as a Gaussian distribution, can be analytically determined. However, for a continuous signal, some error will always be introduced, and the process is referred to as “estimating” rather than “calculating” the PDF. Section 4.3 provides some background discussion of the matter.

All PDF estimators of continuous variables involve some discretisation step, with the kernel and k - Nearest Neighbour (k -NN) methods being common. Kernel estimators rely on counting the number of occurrences of values in specified regions and assumptions about the shape of a signal's probability distribution influence the choice of kernel shape.

A box-kernel estimation method for estimating univariate PDFs works by binning the signal and counting the occurrences of values in specific bins, essentially producing a histogram. The bin width is the bandwidth of the estimator.

k -NN estimators are more computationally efficient, especially for data sparsely distributed over the available dimensions. Since Transfer Entropy (TE) calculation requires estimating PDFs in many dimensions, data sparsity is a real concern.

Java Information Dynamics Toolkit (JIDT) provides options for box-kernel, Kraskov k -NN as well as a few other PDF estimators of continuous signals. The accompanying software integrates the methods above, and Chapter 8 provides comparative studies.

2.6 Caveats

2.6.1 Noise

Excessive measurement noise, if not considered a fault itself, will contribute to the observed novel information in a measurement and compete with the signatures left by real process related faults.

It might, therefore, be desirable to perform rigorous state estimation on all sampled variables and only proceed with the information decomposition tasks on these filtered signals.

Subspace modelling and use of the Extended Kalman Filter (EKF) are potential candidates for a noise reduction preprocessing step, but implementing automated noise handling is not attempted in the scope of this work.

2.6.2 Scaling

In the proposed framework, the amount of uncertainty introduced in the system that impacts significant variables determines the perceived severity of a fault or disturbance. In the ideal case, the value of entropy measured on any signal should have the same meaning between any two variable pairs, regardless of the engineering units used to record the sampled the measurements.

The entropy value of a signal is dependent on the relationship between its variance and the estimator bandwidth used estimate the underlying PDF. To ensure that the entropy of a signal has a consistent meaning throughout an analysis, the scaling of data before estimation of entropy is critical.

The scaling problem is also relevant to other FDD methods such as Principal Component Analysis (PCA) and Reconstruction Based Components (RBC). In the absence of thoughtful preprocessing, most of these techniques only work well when the process has been suitably excited in all dimensions of interest. It is common to employ variance-based dimension reduction techniques on problems where similar processes generated the sampled signals measured in the same units. However, this is rarely the case in chemical process engineering problems.

The variance of standardised signals is unlikely to offer a robust description of unknown interactions in a process engineering environment. A tightly controlled but noisy value will have variance, and thus entropy, blown out of proportion with regards to the real objectives measured. For example, consider a noisy level measurement of 50% with high energy variance of 0.1%. If our control limits for this level (an inventory variable) is between 30% and 80%, this variance does not indicate anything of interest from a process control or monitoring perspective. However, if this signal was standardised, it will likely resemble white noise, which can be easily projected onto almost any other signal in the system to some degree, but will fail thorough significance tests.

A scaling method proposed by Skogestad and Postlethwaite (2005) helps to keep variance in check with regards to the real effects of interest and involves dividing signals by their maximum expected or allowed change, typically the difference between a nominal setpoint and closest limit of interest for a Controlled Variable (CV). Section 4.8 presents a detailed discussion.

2.6.3 Sampling frequency

The processes of a typical chemical plant typically exhibit time constants varying over several orders of magnitude. In the proposed data-driven framework, sampled plant signals are the basis of the PDF estimation step, which in turn are the basis for the calculation of transfer entropy. Therefore, it is essential that the sampled signals capture all dynamics of interest and provide sufficient density in the critical regions of the PDFs.

A complex interaction of factors determines the optimal sampling frequency and Section 9.3.2 discusses this in more detail.

2.6.4 Differentiating between disturbances and faults

The proposed method will indicate process elements where an unknown source is injecting a significant amount of uncertainty the process. Distinguishing whether the novel information is originating from a unmeasured disturbance or fault might be non-trivial in some cases.

Disturbances are more likely to affect multiple nodes simultaneously than a fault originating on a single control or process equipment element, and implementing a set of known disturbance models to condition on could be used to distinguish from disturbances seen before and new process faults.

2.6.5 Differentiating node importance

Not all the states are equally crucial in maintaining optimal operation. Strict control of all process states, such as inventory variables, is not required and might be detrimental to disturbance rejection.

The obtained ranking scores of nodes might require weighting based on their perceived economic or safety importance to refine results and accentuate events of genuine interest.

2.7 Motivation of selected approach

As stated in Chapter 1, the goal of this work is to provide a triaging metric for drawing attention to the most significant process faults at workable computational cost. This dissertation will, therefore, focus on methods that operate in the third degree of simplification from base reality mentioned in Section 2.3.

Based on the views and understanding of the nature and effect of process faults on the information content of signals, and considering the computational constraints and reasonable expectation of search space reduction instead of a full causal explanation, the work of this dissertation will investigate the properties and usefulness of the following two-step process. First, weights are assigned to a directed graph using a bivariate bi-directionally delay optimised directional transfer entropy measure as described in Chapter 4. Next, nodes are ranked using importance scores determined with an inverted eigenvector centrality method as described in Chapter 5.

CHAPTER 3

STRUCTURAL MODELLING OF PROCESSES USING GRAPHS

This chapter discusses established literature regarding structural modelling of process plants using directed graphs and its integration into fault detection and diagnosis methods.

In the context of this dissertation, the term graph refers a collection of nodes and edges (also called arcs). The nodes represent process variables. In a data-driven context, these will be sensors associated with data tags referencing a set of time series data, typically stored in a historian system. Edges represent the relationships that exist between the nodes. Different kind of graphs can be constructed depending on edge properties. The three most common properties assigned to edges are:

Sign

Signed edges can be used to indicate whether there is a positive or negative influence (or process gain) between the nodes, while unsigned edges provide no such information.

Direction

Directed edges indicate the direction of interaction from a source to a sink node. Undirected edges just show the existence of a connection but do not provide any additional information about the relationship between them. A graph with directed edges is called a directed graph or digraph for short.

Edge weights

Weighted edges can be used to indicate the strength of the interaction associated with the edge, whereas un-weighted edges all carry equal importance. In advanced applications, the weight can be a vector of values, each related to a different aspect of interaction strength.

Parameters associated with nodes are used to reflect properties of the modelled processes. In the context of process control engineering, labels could be used to indicate whether the association of a node with a control or non-control process element and provide information on the node's importance from economic and safety perspectives. It could also be used to provide values of KPIs associated with the specific element. Three categories of nodes are (Maurya, 2003*b*):

Exogenous or disturbance variables

Exogenous variables are not affected by any other variable (no incident edges), can change independently and represent disturbance and fault variables.

System variables

System variables are affected by exogenous variables and each other. They may also be referred to as state variables and have both input and output edges associated with them.

Output variables

Output variables do not affect any other variable (no outgoing edges).

3.1 Applications in process engineering

Signed digraphs have been used to model chemical processes for more than thirty years, starting with the work of Iri et al. (1979). Vianna and McGreavy (1995) published the first known application that extended this approach by including weights based on sensitivity coefficients. This idea has seen numerous applications over the last decade in FDD techniques. Signed Directed Graph (SDG) based fault diagnosis methods are popular as they provide a powerful representation to capture cause and effect information about a process, which lends itself to easy exploitation in diagnostic reasoning. SDG models do not require a complete quantitative description of a process as their development makes use of partial information and data-driven methods, as well as heuristics captured from operational experience.

Semi-qualitative analysis of faults is possible via the combination of SDG models with other approaches. Maurya et al. (2006) reported on the integration of SDGs with:

- temporal evolution information,
- weighted graphs and trees,
- multivariate statistics,
- fuzzy logic, and
- Qualitative Trend Analysis (QTA).

Maurya et al. (2007) demonstrated the power of combining SDG with QTA for incipient (early) fault diagnosis, detecting faults in as little as a fifth of the time constant associated with the process. Gao et al. (2010) reported additional work making use of SDGs and QTA to model possible interaction gain directions for fault detection qualitatively.

Chen and Howell (2001) performed extensive work using qualitative reasoning and cause-effect knowledge to assist in isolating faults. According to the proposed scheme, FDD tasks are distributed to separate modules associated with the different control systems located throughout a plant, following a divide and conquer approach.

Harel (1988) introduced higraphs – a combination of conventional graphs and Venn diagrams. Rodriguez and Sanz (2009) presented a combined functional and structural framework of modelling processes as an extension of higraphs called D-higraphs. Later, Díaz and Rodríguez (2014) combined D-higraphs with dynamic simulations of processes for on-line fault diagnosis. Rodriguez and Sanz (2009) also applied these functional modelling techniques to assist with process hazard analysis (HAZOP studies). The principles of higraphs are somewhat different from that for conventional digraph models, and they are not studied further in the scope of this work.

Purely data-driven approaches are found to be lacking when unknown faults or multiple faults are occurring. Chiang and Braatz (2003) proposed incorporating a causal map with data-driven approaches by making use of modified distance and causal dependency metrics to increase fault detection ability. Faulty sensors are detected by the modified distance metric, while causal dependency is used to detect broken dependencies (based on the typical operating conditions). This method relies on analysing deviations in the entropies of the network from the normal operating base case conditions and therefore has the drawback of requiring historical data associated with normal operating conditions, which will require initial effort to catalogue from historian records.

Castaño Arranz and Birk (2012) developed a tool that relies on digraph models of chemical processing plants to assist in the selection of control structures in scenarios involving numerous interactions and variables. Yim et al. (2006) demonstrated the detection of that manual, ratio, feedforward and cascade control configurations via rules exploiting the integration of data-driven analysis with plant topology.

Weidl et al. (2005) used generic object-oriented Bayesian networks, which is a form of Directed Acyclic Graph (DAG) combined with conditional probability distributions, to

model a process which was then successfully deployed in a decision-theoretic troubleshooting and risk assessment application. One of the critical strengths of Bayesian networks is their ability to adapt to changes, such as corrective action taken based on past recommendations.

Note that in all the applications mentioned above the intended use was to help engineers form and test hypotheses. In general, the number of spurious results is too many for unsupervised methods to make final, robust conclusions. The work of this dissertation is no exception, and its intended use is as a filtering mechanism efficient enough to justify its implementation and integration with troubleshooting workflows.

3.2 Developing digraph models of processes

The process of deriving or capturing a structural model of a process has received significant attention over the last two decades. Approaches include knowledge-based, data-driven and hybrid techniques. Knowledge-based methods for determining causality are limited to generating a binary adjacency matrix. In hybrid approaches, knowledge-based topology information ensures that the network structure is correct while data-driven metrics determine appropriate weights for the connections. (Maurya, 2003a) and (Bauer, 2005) enriched the field of finding digraph models of chemical processes and their application to fault detection and diagnosis with two defining studies and associated publications.

The work of Maurya (2003a) is mostly concerned with modelling digraphs from algebraic and differential equations describing the system. Generating non-causal redundant equations by algebraic manipulation on some of the original equations helps to decrease the number of spurious solutions if these contain a different subset of variables. Data-driven methods, specifically PCA was combined with the knowledge derived structure of the digraph to perform online fault detection and diagnosis. Attaching weights to the edges in the signed digraph allows comparison of the relative order of magnitude of deviations in different system variables and further reduce false results.

In data-driven methods, historical data is analysed using various statistical methods to identify likely causal relationships. Methods reported in literature include partial correlation (Farenzena and Trierweiler, 2009), cross correlation time delay estimation (Bauer and Thornhill, 2008), nearest neighbours (Bauer, 2005) as well as the concept of transfer entropy first introduced by Schreiber (2000) and applied by Bauer et al. (2007). Comparing significance levels to threshold values obtained by analysing surrogate random time series data and selecting a required minimum deviation, usually six sigma, helps to reduce false positives.

(Bauer, 2005) compared the performance of the methods for dealing with features typi-

cally found in process time series data such as dead time, low pass filtering and additive noise. Transfer entropy outperformed the other measures in the low pass filtering tests and gave satisfactory results in other areas, proving more robust than cross-correlation and nearest neighbour approaches. It functions well in the absence of noticeable time delays between variables, relying on a different paradigm for measuring the directionality of a disturbance. Transfer entropy is used to analyse fault propagation in ABB's Loop Performance Monitor. Transfer entropy excels as a means of estimating dead time. Shu and Zhao (2013) proposed modifications for increasing transfer entropy's ability to estimate dead time and reported how the obtained time delays could be used to eliminate redundant connections.

Yim et al. (2006) made use of known process topology (physical and logical connectivity) information contained in P&IDs to enhance data-driven plant-wide disturbance analysis tools in a hybrid approach. The known structure is used to verify that a feasible path between a candidate cause and other locations in the plant where secondary disturbances are detected exist, eliminating potential false positives. The method makes use of a reasoning engine written in Prolog on schematics stored in the Computer Aided Engineering Exchange (CAEX) format. CAEX is a neutral Extensible Markup Language (XML) data format intended to allow for exchange between process engineering and control tools. The XMpLant data format also addresses this need. There are already some commercial software packages available that can be used to make object-oriented representations of processes, including CosmosPT from Innotec and Intools or SmarPlant PID&ID from Intergraph.

Thambirajah et al. (2009) expanded on the work of Yim et al. (2006) by incorporating both forward and backwards tests on topological-derived networks to test hypotheses derived from data based methods, thereby reducing the number of spurious results. Maurya et al. (2006) also proposed forward as well as backwards tests.

3.3 Statistical interaction measures

In addition to cross-correlation, nearest neighbour and transfer entropy already discussed, other interaction measures have also been used to infer connectivity or diagnose faults. This section provides an overview of these methods for the sake of completeness. Most of these methods do not detect true causality in a correct sense of the word. See the excellent discussion of Pearl (2009) with regards to the different attitudes and perspectives on the notion of the ability to rigorously measure causality that exists in the modern scientific community.

The Relative Gain Array (RGA) introduced by Bristol in 1966 was one of the first in-

teraction metrics proposed for assisting with the pairing of controlled and manipulated variables. Initially, the RGA was limited to a study of steady-state interactions, but its use was later expanded to all frequencies to provide a measure known as the dynamic RGA (Skogestad and Postlethwaite, 2005).

PCA is a popular and successful tool for monitoring process variations by extracting characteristic features. However, purely data-driven multivariate statistical methods such as PCA and Partial Least Squares (PLS) are inherently limited when the abnormal situations associated with unknown or multiple faults occur.

Contribution plots are one of the most popular tools for locating the specific variables that are pushing the process parameter statistics (typically obtained via PCA) out of their desired control limits and can be used to isolate the faulty root variables under the assumption that those faulty variables make significant contributions to the out-of-control monitoring statistics. Kourti and MacGregor (1996) published some of the earliest work on the use of contribution plots as a means of detecting faulty variables. Contribution plots tend to suffer from a smearing effect which makes it difficult to see which variables best associates with the cause of the fault, and careful interpretation is needed. However, Westerhuis et al. (2000) showed that even though contribution plots do not necessarily indicate the assignable causes, it indicates the group of variables most associated with the detected events, flagging them for more detailed investigation. Liu and Chen (2014) presented a modified contribution plot method which does not suffer from the smearing effect. Unlike various other data-driven techniques, the proposed method does not require a dataset of known faults, making it readily applicable to most industrial operations. The approach was shown to be useful in detecting the propagation of error by controllers to other variables, as well as detecting the variables involved in multiple sensor faults.

Dunia and Qin (1998) developed a reconstruction-based approach to isolate faulty variables from the subspace of faults by extracting fault directions from faulty data, which has become very popular since. Multivariate statistical analysis, such as PCA or PLS is used to characterise fault directions and subspaces on all the fault measurement data. The original RBC method requires that all data associated with a single fault must share the same faulty variables because PCA or PLS will fail to reflect changes in the underlying process patterns as they take the entire measurement data set of each fault case as a single object. Faults propagated by controllers violate this condition. In reality, process patterns such as fault magnitudes or variable correlations will change over time. Zhao et al. (2012) successfully addressed this concern by proposing a Multiple Time Region (MTR) PCA-based fault reconstruction strategy for fault diagnosis. The method relies on an automatic time region division algorithm that partitions the flawed process into different local regions based on changes in the fault characteristics. The fault feature extraction employed to study how faults evolve with time requires characterising the relationship in

process data generated under normal conditions and under the influence of faults, which in turn requires a library of known faults. If such a library did not already exist, it would first need to be built over time via fault detection methods, and only then similar faults will be able to be diagnosed in future. Case studies on the well-known Tennessee Eastman process (Downs and Vogel, 1993) indicated that the MTR fault diagnosis method was more efficient than single time region methods using the PCA statistical measure.

Frequency domain methods also have something to offer. A plant-wide spectral analysis can detect measurements having similar spectral features. Measurements whose spectra are similar are likely to be subject to the same disturbance. Power spectra are invariant under changes in the phase of a signal. Therefore, time delays between tags do not affect the analysis.

Additional prior art studies reported on the use of Fischer Discriminant Analysis (FDA) and Support Vector Machines (SVM)s as fault diagnosis techniques on historically measured data (Zhao et al., 2012).

The scope of this work calls for the developed tool be automated, unsupervised and easy to implement, implying the need for a data-driven approach for inferring connectivity and fault interactions. Also, to facilitate the best distinction between node importances weighed edges will be required. Chapter 4 contains a review of the methods with promise for finding a data-driven causality estimate which can be used to model a weighted directed graph. Chapter 5 presents a study of literature regarding the use of weighted edges in the calculation of node importance.

CHAPTER 4

QUANTIFYING CAUSAL CONNECTIONS FROM TIME SERIES DATA

This chapter provides a detailed discussion of methods and procedures investigated for the calculation of suitable weights whereby the causal links between process elements can be quantified.

Achieving the stated outcomes of ranking elements according to their overall influence on a process requires knowing how each element interacts with its local environment. A good ranking depends on an accurate description of the causal connections that exist among process elements.

4.1 The notion of causality

Owing to frequent misapplication, use of the term “causality” in a technical context justifies a short section to clarify its specific use. Conflict arises from the fact that use of the term “causal relationship” is best reserved to describe a necessary interaction, and yet probability theory – which is most useful for describing systems with a lack of regularity – is frequently encountered in this discipline. The notion that “correlation does not imply causality” is understood by any respectable researcher of this field. Correlation measures predictability instead.

Causal relationship modelling typically applies concepts from the fields of mathematical logic, graph theory, Markov models, Bayesian probability, and many more. Many

scientists consider causal inference as one of the most significant problems in statistics (Hlaváčková-Schindler et al., 2007).

Researchers commonly use both deterministic and probabilistic descriptions of causality. The probabilistic notion is more useful for describing practical, real-life systems and was described by Patrick Suppes in 1970 as follows:

Probabilistic notion of causality (Hlaváčková-Schindler et al., 2007)

An event X is a cause to the event Y if (i) X occurs before Y , (ii), the likelihood of X is non-zero, and (iii) likelihood of occurring Y given X is more than the likelihood of Y occurring alone.

Note that this formulation provides no strict rules for what qualifies as an “event” and is therefore somewhat arbitrary in this respect. Simple statistical measures such as cross-correlation are unable to answer the question of causality presented above as doing so requires an asymmetrical measure at the very least.

Clive W.J. Granger, who won the 2003 Nobel Prize in Economy, introduced the application of causality measures into analyses of data observed in consecutive time instants (also referred to as time series data). He defined a restricted sense of causality.

Definition 4.1.1 (Granger causality). Process X_t Granger causes another process Y_t if future values of Y_t can be better predicted using the past values of X_t and Y_t rather than only past values of Y_t .

A linear regression model as presented in Equation 4.1 is the basis of the standard test for Granger causality.

$$Y_t = a_0 + \sum_{k=1}^L b_{1k} Y_{t-k} + \sum_{k=1}^L b_{2k} X_{t-k} + \xi_t \quad (4.1)$$

The null hypothesis is that X_t does not Granger cause Y_t is supported when $b_{2k} = 0$ for $k = 1, \dots, L$. Alternative derivations include the Granger-Sargent, and Granger-Wald tests (Hlaváčková-Schindler et al., 2007).

The critical requirement of Granger causality is that information about a causative factor should be independently unique to that variable. Removal of the factor under consideration from the universe of all possible causative variables should result in declining predictability of the affected variable (Sugihara et al., 2012). This condition is known as the concept of separability and holds for most strongly coupled variables in non-linear systems. However, in deterministic systems with weak or moderate coupling separability is not satisfied because information of the causal variable is redundantly present in the affected element and removal from the universe of all potential causative variables is therefore not possible (Sugihara et al., 2012).

The linear nature of this model is restrictive, and non-linear generalisations followed from various researchers. Some prominent approaches included analysis of mutual nearest neighbours, use of radial basis functions, application of correlation integrals and conditional entropy (Hlaváčková-Schindler et al., 2007).

4.2 Methods for causality estimation

Estimating causality from a finite number of samples, such as time series data recorded at discrete intervals from a plant over a specific window of time, is not trivial. As is the case in many other disciplines dealing with causality estimation, this application aims to not merely detect or verify whether a causal relationship exists between two process elements, but also to quantify the relative strength of this relationship. Recent publications describe statistical, state-space reconstruction and information-theoretic methods that address this specific need.

4.2.1 Distance correlation

Székely et al. (2007) developed the distance correlation measure primarily to provide a more robust test for independence than that provided by traditional correlation measures.

4.2.2 Directed partial correlation

Mader et al. (2010) discusses the use of directed partial correlation as a measure of interdependence in a multivariate system, in this case, estimating connectivity in the human brain from functional magnetic resonance imaging (fMRI) data. The method is based on vector autoregressive processes and is capable of detecting Granger causality in these systems. No prior knowledge of the studied system is needed.

4.2.3 Convergent cross mapping

Cross Convergent Mapping (CCM) was developed by Sugihara et al. (2012) to address the issue of accurate causality detection in state-dependent systems, that is, situations where a causal relationship between two variables might exist for certain states of the system, but not necessarily at all times. In this framework, the extent to which a historical record of the affected variable Y can reliably estimate states of the causal variable X is used as a test for causality.

CCM resolves the Ganger Causality (GC) shortcoming of requiring separability and can distinguish real interaction from simple correlation produced shared driving variables.

CCM can handle cases of uni- and bi-directional causality, such as found in systems with feedback.

4.2.4 Transfer entropy

Numerous researchers developed information-theoretic methods for the analysis of information flow between components of a complex system over the last few decades. One of the most prominent dynamic, non-linear, non-parametric measures developed for measuring information flow between systems is the transfer entropy functional proposed by Schreiber (2000) and independently presented as conditional mutual information by Paluš et al. (2001), with conditioning on the target variable's past. These measures are information-theoretic functionals of PDFs. Transfer entropy adds information about time to Bayesian networks because it tests hypotheses concerning the joint and conditional probabilities of past and current values in a time series (Hlaváčková-Schindler et al., 2007). Most proposed methods involving TE in the context of FDD try to develop an ITN for the express purpose of reasoning based on its structure (Bauer, 2005; Landman et al., 2014; Yang et al., 2014).

It is essential to distinguish between the total entropy of a signal and its differential entropy. Calculating differential entropies from time series data requires special care since their measurement involves inverse units of time (Kantz and Schreiber, 2003). Also, it is recommended to normalise the result by division with the time lag if a time lag larger than unity is involved in embedding the time series data according to some model - for example, a first-order autoregressive function. Kaiser and Schreiber (2002) provides further discussion on the properties of transfer entropy on continuous data, including convergence and behaviour under coordinate changes.

Convergence of continuous entropy estimates is impossible in most cases. Kaiser and Schreiber (2002) suggests that intuitive interpretations such as “common information content”, “amount of information transferred” or “direction of information flow” be used with extreme caution. These concepts should not depend on the absolute or relative scales or bandwidths used in the analysis, as is the unfortunate case for coarse-grained entropies using the current state of the art methods.

Lizier et al. (2011) made exciting advances in multi-variable causal estimation by developing methods capable of measuring the combined causal relationship of several sources on one or more destination variables. This technique is useful in processes where synergies exist and holds promise for the process engineering context where systems with strong directionality are found. Directionality refers to the idea that a relative movement of two or more variables in specific directions can have a much stronger effect on a destination element than in other directions or the change of a variable in isolation.

The field of process engineering involves highly interactive systems where a significant number of indirect effects can occur. Higher-order connections will confound analysis if included in the result set.

Partial Transfer Entropy (PTE), described by Vakorin et al. (2009), is capable of quantifying indirect coupling caused by the influence of surrounding variables. It differs from the standard transfer entropy calculation by conditioning the measure on all other variables that are indicated to exert an influence on the causal variable under study. Unfortunately, it is not able to deal with bi-directional influences which are common in the process engineering field, partly due to the presence of feedback mechanisms, either inherent in the process itself or via the applied control systems.

Duan et al. (2013) presented the concept of Directed Transfer Entropy (DTE) as a means of indicating causality exclusively between variable pairs not connected via any intermediate or confounding variables. In the ideal case, DTE returns no unwanted higher degree-of-separation connections. The publication claims that DTE is one of the first metrics that allow for direct causality detection from process data for non-linear systems. Calculation of DTE follows standard transfer entropy calculation as a post-processing step to refine connections. An investigation focused on determining appropriate thresholds indicating significant causality for TE, and normalised DTE was reported as ongoing.

Singh and Lesica (2010) introduced the concept of incremental mutual information as an improvement over alternative methods for analysing the dependency among neuron firings. The uncertainty of a variable is first reduced as much as possible by conditioning over past and future values for all variables except the one of interest. The final reduction of entropy by observing the additional variable's value at a specified delay is then measured and reported as the incremental mutual information. The original discussion only concerned binary data streams but showed a reduction in generation of spurious results.

Most proposed methods involving TE in the context of FDD try to develop an ITN for the express purpose of reasoning based on its structure Bauer (2005); Landman et al. (2014); Yang et al. (2014).

4.3 Estimating probability density distributions

A discussion of dynamic non-linear information-theoretic measures of causality is not complete without discussing the means of estimating PDFs from time series data. An ill-advised approach to PDF estimation can have a significant influence on the accuracy of information flow measures.

Histogram and kernel-based methods are two basic ways to estimating PDFs. Histograms are simple nonparametric density estimators defined by the size and endpoints of bins.

Bandwidth is an alternative reference to bin sizes and refers to the level of fidelity retained in the PDF. Histograms are not smooth, and selection of the endpoints can have a significant impact on results if the bandwidth is relatively large.

Kernel functions are a generalisation of, and improvement over histograms. Typical kernel shapes include the uniform (or box), triangular, Epanechnikov (which is optimal from a minimal variance sense), quartic (also referred to as bi-weight), tri-weight, tricube, Gaussian, cosine, logistic and the Silverman kernel. Historically, kernels were hand-picked according to the specific application and nature of data. Duvenaud (2014) published recent developments in automatic selection for Gaussian processes by maximising the approximate marginal likelihood.

Most kernel applications assume that the data can be efficiently modelled by a Gaussian process. A mean and covariance function completely specifies Gaussian processes and can be used to describe time series data from various sources with reasonable success. Assuming a Gaussian process has many significant advantages, mostly related to the ease of integration and closed-form predictive distributions. However, in extreme cases, approaching a non-Gaussian distribution as a Gaussian can result in entropy minimisation instead of maximisation (Hlaváčková-Schindler et al., 2007). Many of the familiar concepts in statistical analysis, such as linear regression and Kalman filters, are examples of kernels which assume a Gaussian process.

Lizier et al. (2008) used box-kernel (uniform kernel) estimation together with dynamic correlation exclusion while optimising box counting for computing differential transfer entropy. Box kernels are discontinuous with a square shape and are similar to histograms in univariate cases. The measure centres a box over each data point, and the overlap of boxes determine the density estimate at each point. The final results are, therefore, non-smooth.

Box kernel PDFs make no assumptions about the nature of distributions (for example, Gaussian), making this method robust to a wide variety of datasets from different sources. It may, however, require more samples to be as accurate as a kernel with a shape that does in fact correctly represent the distribution. As robustness is a more significant concern than computational efficiency in the current implementation, selecting the best kernel shape for each data set encountered is not considered to be a priority. Besides, there are no automated routines for estimating the best kernel choice for non-Gaussian distributions, making the use of individually tuned kernels impractical.

Non-parametric entropy estimators avoid the need for constructing a PDF and integrating over it as is achieved in kernel methods (Stegg, 2013). Instead, the approach estimates a local log probability density at each sample point which is averaged to get the final result. In contrast to kernel density estimates, this procedure is consistent irrespective of the number of samples. An implementation of these methods is available in the Non-

parametric Entropy Estimation Toolbox (NPEET) (Steege, 2013). (Duan et al., 2015) discusses a parameter-free zero information calculation to TE which provides a robust estimator applicable to non-stationary processes but tends to be overly conservative with a high false negative rate.

The k -NN based PDF estimator by Kraskov et al. (2004) is considered the current best-of-breed empirical (conditional) mutual information estimator (Wibral et al., 2014) and is recommended for use in the estimation of transfer entropy (Lizier, 2014).

It is not necessary to estimate transfer entropy on continuous signals for successful use in FDD on chemical processes. The required computational load can be significantly reduced by the use of symbolic encodings (Staniek and Lehnertz, 2008) and discrete approximations of signals based on alarm limits (Yu and Yang, 2015; Su et al., 2017).

4.4 Available software packages

The results presented in this work relied on the implementation of transfer entropy calculators provided by Lizier (2014) which is made available in JIDT.

In principle, form-specific estimators are not needed to calculate the different kind of entropies that can potentially be of interest to use as information-theoretic measures, as they are all combinations of the primary entropy measure. For example, Conditional Mutual Information (CMI), which is identical to TE for our purposes, can be written as a combination of standard entropies as indicated in Equation 4.2:

$$I(X : Y|Z) = H(X, Z) + H(Y, Z) - H(X, Y, Z) - H(Z) \quad (4.2)$$

where $I(X : Y|Z)$ is the conditional mutual information from X to Y , given that Z is known, while $H(X, Y, Z)$, etc. are (multivariate) Shannon entropies.

The danger with this approach to implementation is that the conditional mutual information might be low even though some of the entropies used in the sum can be quite high, and therefore small errors in the estimators can lead to significant errors in the final result (Steege, 2013).

Apart from box-kernel estimation, JIDT also allows for the calculation of transfer entropy according to Kraskov-Stügbauer-Grassberger (KSG) estimators (Kraskov et al., 2004) of mutual information. KSG estimators make use of k -nearest neighbour statistics and were shown to provide a significant improvement when deployed in Independent Component Analysis (ICA) against other methods, especially with regards to quantifying the independence among the components. Mutual information estimated according to KSG estimators are exact for independent distributions, returning the expected zero value up

to statistical fluctuations. This work compares the performance of box-kernel estimators with KSG estimators for the particular application at hand.

Note that JIDT is still under active development. Future researchers will do well to make themselves aware of the new measures and estimators available.

Another package related to entropy estimation is the THOTH package implemented in Python (DeDeo et al., 2013), which focuses on entropy estimators on under-sampled discrete data using Bayesian and bootstrap estimators (Nemenman et al., 2004). This approach might be useful if data from composition analysers or other sensors with prolonged sampling intervals are involved but is not investigated further in the scope of this work.

4.5 Choice of methods

Bauer et al. (2005) investigated the usefulness of cross-correlation, nearest neighbour methods as well as transfer entropy for inferring connectivity from time series data recorded from chemical processing plants. Bauer et al. (2007) discusses the use of transfer entropy for finding the direction of disturbance propagation in a chemical process and also demonstrates the creation of directed graph models using transfer entropy to infer connections. These publications motivated the choice for selecting transfer entropy as one the main metrics to investigate for use in addressing the ranking problem.

As cross-correlation is much more commonly known and significantly less computationally intensive, it was decided to compare its usefulness with that of transfer entropy, even though it is not a measure of causality. Cross-correlation does excel in specific areas where transfer entropy fails, such as detecting similarity between signals, and therefore it is possible that a hybrid method might be the best solution.

Details with regards to cross-correlation and transfer entropy calculation, significance testing and optimal parameters for dealing with process data, especially when the desire is not merely to detect but also to quantify causality, is discussed in Section 4.6.

The ranking method developed in this work desires quantification of the relative sizes of the entropies calculated between process elements. Fixing the bandwidth and sampling interval for a specific set of calculations leaves the lag associated with the processes itself as a significant variable in need of consideration. Chapter 6 studies the relationship between sampling interval, dominant process time constant and the absolute value of the entropies calculated.

4.6 Formal definitions of measures used

4.6.1 Transfer entropy

Schreiber (2000) originally defined transfer entropy as follows,

$$T_{Y \rightarrow X} = \sum_{(x_{i+h}, \mathbf{x}_i^k, \mathbf{y}_i^l)} p(x_{i+1}, \mathbf{x}_i^k, \mathbf{y}_i^l) \log \frac{p(x_{i+1} | \mathbf{x}_i^k, \mathbf{y}_i^l)}{p(x_{i+1} | \mathbf{x}_i^k)}, \quad (4.3)$$

where Y is the source and X is the target variable, making use of l and k historical source and target values respectively.

The values of k and l is known as embedding dimensions. In the most naive implementations, these are both set to unity. However, to prevent over-estimation of the transfer entropy, self-information needs to be maximised in the embedding according to some criteria. With auto-embedding enabled for the Kraskov estimator, the JIDT algorithm performs this optimisation according to the Ragwitz criteria. An alternative optimisation involves the maximisation of Active Information Storage (AIS).

Bauer (2005) desired to measure transfer entropy at various time delays, proposing the following formulation

$$T_{Y \rightarrow X} = \sum_{(x_{i+h}, \mathbf{x}_i^k, \mathbf{y}_i^l)} p(x_{i+h}, \mathbf{x}_i^k, \mathbf{y}_i^l) \log \frac{p(x_{i+h} | \mathbf{x}_i^k, \mathbf{y}_i^l)}{p(x_{i+h} | \mathbf{x}_i^k)}, \quad (4.4)$$

where h indicates the time delay in sampling intervals (also known as the prediction horizon).

Shu and Zhao (2013) showed that the following definition is more representative of what is actually attempted

$$T_{Y \rightarrow X} = \sum_{(x_{i+h}, \mathbf{x}_{i+h-1}^k, \mathbf{y}_i^l)} p(x_{i+h}, \mathbf{x}_{i+h-1}^k, \mathbf{y}_i^l) \log \frac{p(x_{i+h} | \mathbf{x}_{i+h-1}^k, \mathbf{y}_i^l)}{p(x_{i+h} | \mathbf{x}_{i+h-1}^k)}. \quad (4.5)$$

This formulation fixes the reference, allowing estimation of the time delay by maximising transfer entropy over a suitable range of prediction horizons. The calculation according to Equation 4.5 is easily achieved by shifting the source and sink vectors on each other before calculation according to standard implementations.

Lizier et al. (2008) showed that the estimate could be made significantly less computationally intensive by re-formulating the equation as a global average of local transfer entropies and taking into account that the calculation involves a finite number of observations. This formulation efficiently eliminates one PDF estimation and significantly

reduces the number of states summed by the algorithm. The formulation according to this simplification with time delays incorporated is as follows

$$T_{Y \rightarrow X} = \frac{1}{N} \sum_{i=1}^N \log_2 \frac{p(x_{i+h} | \mathbf{x}_{i+h-1}^k, \mathbf{y}_i^l)}{p(x_{i+h} | \mathbf{x}_{i+h-1}^k)}. \quad (4.6)$$

All entropy calculations reported in this dissertation make use of base 2 logarithms. The unit of the entropy results will, therefore, be in bits.

The transfer entropy calculator implemented in JIDT that makes use of the Kraskov estimation method reports values in nats (base e logarithm). Nats can be converted to bits simply by multiplying by $\log_2(e)$ according to the following logarithmic conversion rule:

$$\log_e(x) = \frac{\log_2(x)}{\log_2(e)} \quad (4.7)$$

Bauer (2005) utilised the difference between forward and backwards transfer entropies are to maximise the potential benefit obtained from the directionality of transfer entropy as shown in Equation 4.8. This option is available in the accompanying software and presented in most results in this work.

$$t_{Y \rightarrow X} = T_{Y \rightarrow X} - T_{X \rightarrow Y} \quad (4.8)$$

4.6.2 Cross-correlation

Cross-correlation is a widely known and applied statistical measure. Although correlation is necessary for causality, correlation does not guarantee it.¹ This section includes discussion of cross-correlation primarily for comparison to the information-theoretic measure of transfer entropy.

The linear correlation coefficient, also referred to as Pearson's correlation coefficient, is defined as

$$r_{xy} = \frac{\sigma_{xy}^2}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (4.9)$$

where σ_{xy}^2 is the covariance of the two signals and σ_x and σ_y the standard deviations. The general cross-correlation function for two vectors lagged by a discrete time interval κ is

$$\phi_{xy}[\kappa] = \frac{1}{N} \sum_{i=-N}^{+N} x_i y_{i-\kappa}. \quad (4.10)$$

¹For a detailed discussion of the issues surrounding correlation and causality inference, see Section 4.1

If the input signals are mean-centred and the correlation coefficients standardised to unit variance, the cross-correlation function returns an identical result as the linear correlation coefficient with one variable shifted by a specific number of discrete samples κ .

Scaling variable variance according to the desired control limits is important for the weights between variables to indicate a relative importance (see Section 4.8). As Pearson's correlation coefficient inherently standardises the signals, the suggested method uses the general cross-correlation coefficient.

4.7 Significance testing

In a fully data-driven scheme for fault and disturbance screening, no *a priori* knowledge is used to force any meaningful structure to the results obtained, making it crucial to indicate relative confidence in results. The suggested analysis workflow incorporates significance testing on the cross-correlations as well as transfer entropy estimates in an attempt to detect whether a calculated value represents a real or spurious interaction.

4.7.1 Transfer entropy

4.7.1.1 Magnitude

Schreiber and Schmitz (2000) discusses various aspects of testing whether a non-linear description of data signals is justified and of practical importance. Empirical studies showed that non-linear measures do not follow a Gaussian distribution in general. For this reason, Schreiber and Schmitz (2000) discourages the somewhat common practice of representing the distribution of a non-linearity metric by an error bar and deriving the significance of the calculated values from the number of standard deviations the data lies outside these bounds. Bauer (2005) proposed that a transfer entropy value indicate a significant relationship if the difference between the transfer entropy calculated for the original data and the mean transfer entropy for a set of appropriately generated surrogate data sets is more than six standard deviations away from the standard deviation of the set of surrogate transfer entropies. The six-sigma criterion is selected to compensate for the fact that the transfer entropies do not follow a Gaussian distribution. The significance threshold for transfer entropy from Y to X , $s_{x \rightarrow y}$, is given by

$$s_{x \rightarrow y} = \frac{\lambda_0 - \mu_\lambda}{\sigma_\lambda} > 6 \quad (4.11)$$

where λ_0 is the transfer entropy $t_{x \rightarrow y}$ (the null hypothesis of no significant magnitude) while μ_λ and σ_λ is the mean and standard deviation of the transfer entropies calculated

on the set of surrogate data.

This work primarily follows the suggestion of Schreiber and Schmitz (2000) by using a non-parametric rank-order method. According to a rank-order test, a residual probability of a false rejection α corresponds to a significance level of $(1 - \alpha) \times 100\%$. A test for significance in a specific direction of time is one-sided. For one-sided tests, $M = 1/\alpha - 1$ surrogate sequences are generated, giving a total of $1/\alpha$ data sets including the data set itself. The probability that the data gives the most significant causality measure value by coincidence is, therefore, α . For a significance level of 95%, the test thus generates and calculates the transfer entropy of 19 surrogate data sets and requires that the real data returns results higher than all the surrogates to pass.

This significance testing approach is computationally expensive. The software provides an option for performing the significance test only after maximising the measure over a range of time delays. However, for detailed analysis, it is still possible to request significance to be tested at every delay in the optimisation search space.

The kind of surrogate data used to compare the data with plays a role in what differences are analysed and will, therefore, affect the results of the tests for different processes. For example, a significance test making use of white noise will only confirm whether the relationship between the source and destination time series data is better than the causal relationship between white noise and the target variable, a relatively low bar.

Bauer (2005) advocates the calculation of surrogate data according to the iterative Amplitude Adjusted Fourier Transform (iAAFT) method (Schreiber and Schmitz, 2000) whereby the surrogate data is iteratively adjusted to match both the amplitude distribution as well as the power spectrum of the original data.

Marschinski and Kantz (2002) defined effective transfer entropy as the difference between transfer entropy calculated from the real data set and the source data shuffled in time as shown in Equation 4.12.

$$ET(X|Y) = T(X|Y) - T(X|Y^{\text{surr}}) \quad (4.12)$$

Lizier et al. (2011) also made use of random shuffling to generate surrogate data for significance testing. Both iAAFT and random shuffle surrogate data generating methods are available as options for testing significance according to the rank-order method in the accompanying software.

4.7.1.2 Directionality

Although transfer entropy, and especially directional transfer entropy as defined in Equation 4.8) is an asymmetric measure and therefore implicitly tests significance in a specific

direction, optimisation over a range of delays complicates matters as targets might appear to influence sources if the source value is lagged far enough behind the target value.

An optional directionality significance test is available to help reduce spurious connections related to this effect. The test involves calculating the transfer entropy over a range of delays shifting the source target forward and backward, and recording the maximum values and their associated delays in each direction, $[\psi^{\text{forward}}; \delta^{\text{forward}}]$ and $[\psi^{\text{backward}}; \delta^{\text{backward}}]$ respectively.

The directionality test passes on condition that $\psi^{\text{backward}} > \psi^{\text{forward}}$ or $\delta^{\text{backward}} < \delta^{\text{forward}}$.

4.7.2 Cross-correlation

Bauer (2005) developed a significance test for cross-correlation. It consists of both an absolute magnitude and directionality threshold, both which needs to be passed for the measured value to be considered indicative of a significant causal relationship. The significance test used in this work is based on the same concepts, but adjusted for the fact that the general cross-correlation coefficient is used instead of Pearson's correlation coefficient.

The first step in testing the significance according to the absolute strength of the measure is to optimise the correlation value over a range of potential time delays. Since correlation values can either be positive or negative and the sign does not indicate directionality as for the case of directional transfer entropy, the algorithm estimates the time delay by searching for the most substantial *absolute* magnitude linear correlation over a suitable range of delays. It is important to note that, unlike the case of transfer entropy, the range of time delays extends equally in both directions.

A positive optimal time delay will imply that x causes y , as even according to the most basic definitions of causality the cause must precede the effect in time. However, negative time delays imply that y causes x and in this case the algorithm will set the inferred value between the source and variable target to zero. Since a fully connected matrix is tested, this connection should be picked up when the source and target order are reversed.

Cross-correlation is a symmetrical measure and therefore an additional directionality test is needed to confirm the significance of the calculated value for a specific source-target bivariate pair of variables. Consider the case of time-reversal periodic (symmetrical with respect to time) oscillatory signals. If the oscillation period is T_{osc} , the absolute value of the correlation measures optimized by a time delay λ will be similar that obtained by a time delay of $\lambda - \frac{T_{\text{osc}}}{2}$, which may be negative.

Therefore, to argue for a cause and effect relationship using a correlation measure, a directionality test needs to be implemented at the expense of not being able to deal very well with bi-directed coupled systems. The directionality test proposed by Bauer (2005)

asserts whether there is a significant difference between the maximum and minimum correlation value as follows:

$$\omega = \frac{|\phi^{\max} + \phi^{\min}|}{\frac{1}{2}(\phi^{\max} + |\phi^{\min}|)} \quad (4.13)$$

where ϕ^{\max} and ϕ^{\min} is the maximum and minimum linear correlation values obtained over the range of delays investigated. Since ϕ in this case is the linear correlation coefficient and as such is limited to $-1 \leq \phi \leq 1$, and ϕ^{\max} and ϕ^{\min} are expected to have opposite sign, the possible range of the directionality index is $0 \leq \omega \leq 2$, with larger values indicating more distinct magnitudes.

Bauer and Thornhill (2008) determined threshold values for both the absolute magnitude as well as the oscillation index based on results obtained between a thousand different uncorrelated random sequences. The threshold depends on the sample lengths involved. They recommend using an absolute linear correlation measure threshold of the mean of the three standard deviations. Based on an empirical study it was determined that the appropriate correlation threshold values are (Bauer and Thornhill, 2008)

$$r_{\text{th}}(N) = 1.85N^{-0.41} + 2.37N^{-0.53} \quad (4.14)$$

where $N \geq 200$.

A similar approach was followed to determine threshold values for the directionality index ω according to a one-sided test with a peak close to zero. Bauer and Thornhill (2008) used a single standard deviation from the zero value as the threshold, rather than the more common three standard deviations, motivated by the fact that the test missed many relationships (produced false negatives) if three standard deviations were required to indicate significance. The corresponding empirical fit for the directionality index fit is as follows:

$$\omega_{\text{th}} = 0.46N^{-0.16}. \quad (4.15)$$

As general cross-correlation is used instead of Pearson's correlation coefficient, two adjustments to the significance testing strategy had to be made.

As the cross-correlation values are heavily dependent on the mean and variance of the signals which are no longer standardised, no general threshold values based on once-off empirical studies can be obtained. This problem is addressed by performing an empirical study on surrogate data for every pair of data in similar fashion to the case of transfer entropy described in Section 4.7.1.

In addition, the directionality index defined in 4.13 relies on the assumption that ϕ^{\min} and ϕ^{\max} is guaranteed to be negative and positive, respectively. This requirement is frequently violated by the general cross-correlation measure, particularly for slow processes

where the delay range tested might be insufficient to capture the full range of dynamic effects.

This problem can be overcome by defining the directionality index:

$$\omega = \frac{|\phi^{\text{forward}} - \phi^{\text{backward}}|}{\frac{1}{2}(\phi^{\text{forward}} + \phi^{\text{backward}})}, \quad (4.16)$$

where ϕ^{forward} is the maximum absolute value of cross-correlation in the forward direction (where $\kappa \geq 0$) and ϕ^{backward} is the maximum absolute value of cross-correlation in the backward direction (where $\kappa < 0$).

4.8 Scaling

It is important to ensure that the magnitude of calculations are consistent with relevant operational information content and are independent of engineering units used for recording. It is common practice to standardise (mean centring and unit variance scaling) data before performing analysis steps (such as PCA). This scaling approach introduces the requirement that sufficient and representative excitation of the process in all directions should be present in the sample data. If this is not the case, dividing by a small standard deviation blows a small amount of noise on otherwise quiet signals out of proportion, misrepresenting the relative real information content carried by the signals.

A key desired property of the method developed in this work is that it requires no particular process conditions excitations to be useful, and as such standardisation of data is not a suitable strategy for dealing with the data scaling problem.

Shifting the data by a constant does not affect either the cross-correlation or transfer entropy calculations while scaling the data has a drastic influence on the calculated values for transfer entropy, with values varying with several orders of magnitude.

Standardisation removes all information about how much the signal varies relative to its desired operating range. For example, there is no need to flag a temperature value that is influenced by some element to vary between 349 and 351 K while the desired control limits are 330-370 K as a critical causal link. On the other hand, that same disturbance influencing an analyser composition reading to fluctuate between 0.85 and 0.86 while the minimum requirement is 0.84 should be regarded as highly influential. There is, therefore, a clear need for introducing some means of normalising weights based on how significant the standard deviation of the original time signals are relative to the nominal value and the desired control limits. If the variable under consideration is a manipulated variable, it is necessary to take the available range relative to the current operating range

into account. Data scaling guidelines proposed by Skogestad and Postlethwaite (2005) addresses both of these concerns.

The scaling scheme of Skogestad and Postlethwaite (2005) subtracts the nominal operating value from the value of each variable, and the result is divided by its maximum expected or allowed change. The result is that all values are scaled to have a value less than 1 in magnitude as long as they remain within the predetermined limits, ensuring that all variables are of equal importance regarding their magnitude. If the nominal operational point does not lie in the middle of the allowed or desired range, the most conservative approach is to use the biggest difference to scale disturbances and reference values, while the smallest difference is used to scale allowed input changes and control errors.

Castaño Arranz and Birk (2012) makes a strong case for the necessity of getting a scaling independent representation of data to make reliable conclusions about relative interaction strengths among process elements using weighted graphs. The author based the suggested means of achieving such a scaling on the same concept also expressed by Skogestad and Postlethwaite (2005). Motivated by the fact that it can be tedious to find appropriate scaling factors for each process measurement according to this scheme, they developed a method that achieves a scale-invariant representation using an estimation of the standard deviation of the signals. To obtain this information the process, or a model thereof, needs to be excited with signals by variances selected according to the expected or allowed upsets in process outputs.

Subjecting processes to excitations of the required magnitude are undesirable in the industry as it can be difficult to prove that obtaining this information will be worth the effect this testing has on product quality and process stability during testing. However, it is not unlikely to envision that future progress in signal processing and decomposition will allow the required variances to be obtained from routine operating data, or from unintuitive automated controlled testing, as was done in recent years for the case of identifying the process models required by MPCs.

CHAPTER 5

RANKING PROCESS ELEMENTS

This chapter describes concepts related to ranking critical control loops in processes by means of graph centrality measures useful for indicating relative importance among elements in a directed graph.

Plant-wide controller performance assessment is a familiar and active field in the control engineering academic society, and has been for many decades (see Thornhill and Horch (2006) for an excellent review). Various researchers, most notably Farenzena et al. (2009); Farenzena and Trierweiler (2009) and Rahman and Choudhury (2011), addressed the concept of ranking faults and control loops based on their interactions relatively recently, and this approach is yet to see widespread industrial use. Many proposed methods require plants to be step tested to determine useful interaction measures. Step testing can be costly and in the case of already unstable plants near impossible. However, in most cases, a step size of 1-2% of the nominal value is sufficient, depending on noise levels (Rahman and Choudhury, 2011).

5.1 Overview of prior art

Rossi et al. (2006) proposed two methods for quantifying loop interaction as well as a performance index for the servo control problem (set-point tracking). Both of these methods require set-point excitations to establish interaction measures. The first method analyses interaction in the frequency domain based on the Power Spectral Correlation Index (PSCI) of the error while the second approach operates in the time domain by evaluation of a modified Integrated Absolute Error (IAE). As oscillatory set-point changes are

not commonplace in chemical industries, the time domain technique might be preferred over the frequency domain one.

Rossi et al. (2006) incorporated these interaction measures into a control loop performance diagnostic method designed to determine whether a decentralised PI(D) controller will be sufficient or if an advanced control structure such as a MPC is justified. The Controller Performance Index (CPI) - determined by comparing the error to the theoretical minimum that could be achieved by a minimum variance controller - is used to assist in making this decision. A CPI close to zero indicates that the performance is near optimum, while a CPI close to one indicates that there is significant room for improvement, with a threshold value of 0.5 proposed. Below this threshold, re-tuning would provide little benefit since the minimum variance controller is an idealistic case and of little practical use.

Determining the CPI requires estimation of a time delay parameter. Rossi et al. (2006) approached this by approximating the open-loop response to a second-order transfer function model and optimising parameters in the least square sense.

The IAE interaction metric employs a normalisation factor to ensure independence from the measurement scale. This measure is similar to that prescribed by Skogestad and Postlethwaite (2005) and is equal to the distance between the set-point mean and closest limit. Use of control limit based scaling instead of standardisation assists in indicating which control loops have more stringent requirements on them, and should therefore also enjoy greater performance priority.

The application of PSCI allows the exclusion of loops characterised by different frequencies due to other oscillating sources. PSCI lies in the range $[0, 1]$ with similar shapes of power spectra giving a more substantial value and indicating the presence of interaction. Error based interaction measures rely on the idea that a change in any parameter of one control loop (such as its set-point) will increase the integrated error of any other loop affected by the adjusted loop. In the modified IAE method proposed, the IAE is set to zero if the error changes sign, allowing the difference between excitations (more massive peaks) and noise (smaller peaks) to be clear. The IAE peaks thus obtained are compared for different variables and directly used to evaluate the interaction between loops.

A time window based approach is defined according to the duration of peaks and helps to ensure grouped analysis of peaks referring to the same excitation. The Interaction in Time Domain (ITD) metric is calculated based on the IAE of the windowed data trend. If the frequency or time domain metrics lies above 0.5, this indicates that set-point change in a specific loop affects other loops to a greater extent than the loop itself.

Rahman and Choudhury (2011) presented the use of two data-driven methods for calculating interaction among loops, namely canonical correlation as well as error measures such as IAE or Integrated Square of Error (ISE) similar to that submitted by Rossi et al.

(2006). The presented method determines a loop interaction index after arranging the measures into a matrix form.

Canonical Correlation Analysis (CCA) can be used to quantify relations between multi-dimensional variables under the assumption that the relationships are linear. The most significant difference between CCA and ordinary correlation analysis is invariance on affine transformations of the variables, making the measure less sensitive to scaling and different offsets between the variables.

Farenzena et al. (2009) suggested prioritising control loop maintenance according to a variability matrix. The variability matrix indicates how improving the performance of a specific loop to the least variability possible can be expected to influence the variability in other control loops, and can be used to observe the best compromise situation. In many cases, reduced variability in one control loop will increase variability in control loops interacting with this loop. However, in cases where interaction helps to propagate a disturbance through a system, reducing the variability in a loop will also result in reduced variability in other loops.

Different methods for determining a variability matrix are suggested, depending on whether a model of the process and controller is available or not. The usefulness of the variability matrix is claimed to be insensitive to model mismatch. As performance is not the key criteria for many loops, especially those associated with buffer elements such as surge tanks, they suggested calculating a corresponding variability matrix which compares the variability of a loop with the worst acceptable variability.

Another important concept presented in this work is the use of a Control Loop Economic Benefit (CLEB) array, which converts the reduced variability in each loop to a value per unit time. The CLEB provides a direct measure of the expected value of minimising the variability of a control loop. Since reducing the variability of a control loop can lead to increased variability in other control loops, it comes as no surprise that increasing the performance of specific loops will have a negative economic impact. In these cases, the CLEB array will correctly have negative entries. The CLEB is a compelling concept, but assigning economic values to variability reduction is far from trivial in industrial plants where the actual values of streams are sometimes far from certain. Farenzena et al. (2009) suggested that the weights used in the optimisation layer employed in a MPC design, if available, be used.

The usefulness of this approach is dubious since a dynamically optimising MPC with weights correctly indicating the real economics should move the variability to those areas where it has the least impact already. However, the integration of the two concepts might result in the potential of developing a method for verifying that the behaviour of an MPC results in economic efficiency.

Farenzena and Trierweiler (2009) were the first to report on the usage of a variant of Google’s PageRank-type algorithm for analysing connectivity between control loops, dubbing the method LoopRank when used in this context. In its simplest form, the LoopRank algorithm is a simple eigenvector centrality measure calculated on a matrix representation of the (weighted) directed graph, see Sections 5.2 and 5.3.4 for details. The idea of combining graph ranking algorithms commonly employed in the fields of computer science and sociology with causal and structural models of process plants served as the primary inspiration for this dissertation, which is an attempt to expand on this concept and study its value.

The presented method relies on the calculation of a relative weight matrix, which is virtually identical to a weighted digraph model of a plant. In the publication mentioned above, the suggested means of determining the weighted digraph model was calculating the partial correlation among the process elements. The ranking results are quite sensitive to the relative weights. Section 4 contains a motivation that explains the need for better causal estimators and reviews alternatives.

The rest of this chapter is devoted to the specifics of determining node importance and interaction according to graph theory analysis.

5.2 Matrix representation of a directed graph

Most graph-theoretic analyses makes use of matrix algebra operators. The digraph can be represented in matrix form as an adjacency matrix,

$$A_{ij} = \begin{cases} 1, & \text{if there is a connection from node } j \text{ directed towards node } i \\ 0, & \text{otherwise} \end{cases} \quad (5.1)$$

as indicated in Figure 5.1. There appears to be no single agreed-upon convention whether source nodes should be along the rows and sink nodes along the columns, or vice versa. This dissertation uses the convention of sources along columns and sinks along rows. In a weighted digraph, the Boolean entries of the adjacency matrix are replaced by the appropriate weights.

5.3 Centrality measures on a directed graph

In graph theory and network analysis, the centrality of a vertex measures its relative importance within a graph. Centrality indices measure the position of a node along a predefined set of walks on the graph, see for example Section 5.4. Various centrality

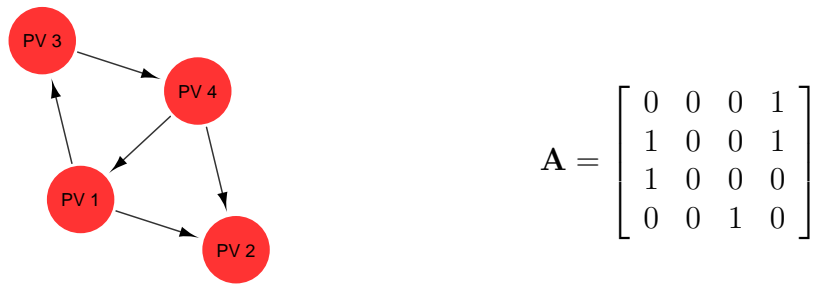


Figure 5.1: Example directed graph with associated adjacency matrix

measures have been defined over time, depending on the sense of importance involved. The following section presents a review of those most relevant to the scope of this work.

5.3.1 Degree centrality

Degree centrality is defined by the number of links incident on a node and relates to the risk of a node of being influenced by whatever is flowing through the network. Directed graph nodes have both an in- and out-degree with defined centrality measures for both. Degree centrality can be used to describe the role of a graph within a graph Zhukov (2015).

5.3.2 Closeness centrality

The length of the shortest paths between a node and all other nodes in a network is a natural distance metric. Closeness is defined as the inverse the sum of all the distances from a node to all other nodes, with lower values indicating a more central node Zhukov (2015).

5.3.3 Betweenness centrality

Betweenness centrality measures the importance of a node by quantifying its total occurrence as an element in the shortest path between any two other nodes Zhukov (2015). Originally introduced as a measure for quantifying the control of a person on the communication between other users in a social network, it might be of particular interest in a chemical process tag network. In this context, a tag that frequently occurs in paths between other identified tags will have a high likelihood of being the cause of a fault or disturbance.

5.3.4 Eigenvector centrality

Eigenvector centrality measures the influence of a node in a network as being equal to the sum of importance scores of all nodes with an edge incident to it, given by

$$C_{\text{eig}}(k) = \sum_{j \in \mathbf{L}_k} w_{kj} x_j \quad (5.2)$$

where $C_{\text{eig}}(k)$ is the importance of node k according to the eigenvector centrality measure, \mathbf{L}_k is the set of nodes which is influenced by x_k and \mathbf{W} is the ranking weight matrix Zhukov (2015).

The entries of the rank weight matrix \mathbf{W} should be real numbers presenting connection strengths as in a stochastic matrix. It should also be a normalised adjacency matrix such that each column sums to unity and all importance scores are positive.

Finding the solution of each node's importance involves solving simultaneous linear equations. The problem can be rewritten in the format of a standard (right) eigenvalue problem given by

$$\mathbf{W}\mathbf{x} = \lambda\mathbf{x}. \quad (5.3)$$

There are many eigenvalues λ with different eigenvectors \mathbf{x} which will satisfy this equation. The eigenvector with all positive entries and associated with the highest eigenvalue contains the relative importance scores and is related to the stationary probability vector defined on a stochastic matrix, that is, a vector that does not change under application of the transition matrix. The Perron-Frobenius theorem guarantees the existence and uniqueness of such an eigenvector. The proof relates to the idea that the long-term probability of being in a state is independent of the initial state. This is discussed in greater depth in Section 5.4. If the weight matrix \mathbf{W} is column-stochastic (all columns sum to one), as is desired, the largest eigenvalue will be unity, $\lambda = 1$.

In the context of this work, the goal of the node ranking algorithm is to make use of a weighted digraph to infer which nodes have the most significant impact on the network as a whole and are therefore likely to play a vital role in introducing or propagating a fault or disturbance in the process. The eigenvector centrality concept is ideal for this purpose and is also a fundamental element of the PageRank algorithm (Bryan and Leise, 2006).

5.3.5 Katz centrality

Katz centrality measures the total number of walks (sequence of edges and nodes connecting two nodes) between two nodes in the network. In effect, it is similar to eigenvector centrality with the additional accounting for higher order connections based on some attenuation factor.

Equation 5.4 provides a formal definition of Katz centrality.

$$C_{\text{Katz}}(k) = \sum_{p=1}^{\infty} \sum_{j=1}^N \alpha^p (W^p)_{kj} \quad (5.4)$$

where $C_{\text{Katz}}(k)$ is the importance of node k according to the Katz centrality measure, α is known as an attenuation factor and \mathbf{W} is the weight matrix. The attenuation factor α cannot exceed a value of $1/\lambda$ where λ is the largest eigenvalue of the ranking weight matrix \mathbf{W} .

Unlike eigenvector centrality, the definition of Katz centrality is valid on DAGs, which removes the need for adjusting the adjacency matrix with a fully connected matrix with low weight (see Section 5.4 for details).

The principal eigenvector associated with the largest eigenvalue of the ranking weight matrix is returned by the Katz centrality measure in the limit that α approaches $1/\lambda$ from below. This is identical to the result calculated according to eigenvector centrality in the limit of weighing an acyclic graph by a uniform reset bias vector with an attenuation factor that approaches zero from above.

This work investigates both eigenvector and Katz centrality. Chapter 7 explores practical considerations and presents sample ranking problems.

5.4 Markov chain interpretation

The idea of assigning importance to a node by analysing the distribution of time a random walk starting on a random node will end up spending on the various nodes is the basis of the core ideas employed in the Google PageRank algorithm, which is similar to the eigenvector centrality measure. A useful interpretation of the eigenvalue centrality measure is that it provides the results of a Markov chain operator on the graph.

The Markov chain problem formulation usually consists of a Transition Probability Matrix (TPM) \mathbf{G} (also called a substitution matrix) together with a Reset Probability Matrix (RPM) \mathbf{Q} weighed by a mixing factor m , where

$$\mathbf{M} = m\mathbf{G} + (1 - m)\mathbf{Q}. \quad (5.5)$$

The RPM indicates the probability that an actor will move randomly to any node on the network. In basic algorithms, this probability distribution is usually considered to be equal among all nodes.

The TPM indicates the chance that an actor currently residing on a specific node will move to a neighbouring node – in the appropriate direction for the case of directed graphs. Since the sum of all probabilities for a given event should always sum to unity, both the TPM and RPM consists of stochastic column vectors containing non-negative real numbers. A vector is stochastic if its elements sum to unity.

For the Markov chain operator interpretation to hold, it is essential that there be no dangling nodes, that is, columns which have no non-zero entries (whose sum is, therefore, zero). As such, in the context of this work, it is unable to compute if there are nodes that do not influence any other nodes, which is very likely to occur in most circumstances. A straightforward means of dealing with dangling nodes is to add a very weak connectivity between all nodes in the network. Weighing the TPM \mathbf{G} with an equal probability RPM \mathbf{Q} according to a weight m very close to unity implements this solution.

In the original PageRank implementation, m was set at 0.85 while \mathbf{Q} was an equally weighted matrix, such that there is a reset probability of going to a random node 15% at any transition according to the Markov chain walk interpretation of the ranking scores. The motivation behind this is not necessarily applicable to the ranking of process elements. If the reset probability matrix is selected to be a simple equally weighted fully connected matrix, the researcher should set m as high as possible while still allowing for a stable numerical solution to be obtained, preventing unnecessary “dilution” of causal link information obtained by causal effect significance estimation. The actual value will depend on the differences in magnitudes for the weights and the number of iterations allowed.

5.5 Incorporating meta-data in the ranking problem

Identifying the most critical faults will be significantly assisted by incorporating individual controller performance KPIs and stream importance measures, such as economic value, as meta-data in the node ranking problem. For example, since the purpose is to accurately identify poorly performing control loops that have the most significant impact on the plant, a higher weight might be assigned to control loops with poor individual control loop KPIs, which are readily available on many plants utilising some form of control performance analysis software.

Adjusting the relative reset probabilities is an ideal means of biasing the network ranking towards specific nodes and can, therefore, be used to incorporate additional meta-data

in the ranking algorithm. Personalised PageRank makes use of a weighed RPM to give preference to nodes indicated as being of particular interest to a specific user based on historical data.

The biasing scheme requires a relative reset bias vector \mathbf{q} to be defined. This vector is normalised such that it sums to unity. Therefore the relative, and not the absolute, values of the entries in the reset bias vector is of importance.

Now, the reset probability matrix \mathbf{Q} is simply defined as the $n \times n$ matrix with each row equal to the normalised reset bias vector \mathbf{q} :

$$\mathbf{Q} = \begin{bmatrix} \mathbf{q}, \mathbf{q}, \dots \end{bmatrix}^{\top}. \quad (5.6)$$

Notice that the reset probability matrix \mathbf{Q} will be row-stochastic, but not necessarily column-stochastic.

The transition to reset probability mixing weight m can be considered a tuning factor that will determine how significant the edge weightings are relative to the node biases in determining the final ranking score if additional information is incorporated.

Section 7.3.2 provides demonstrations of the use of the reset bias vector.

If an application requires the consideration of more complex factors, an alternative approach may be needed. Gao et al. (2011) presented algorithms which allow for optimising a network ranking by considering meta-data associated with nodes as well as edges. In the proposed scheme the ranking problem is solved in conjunction with some objective functions in a semi-supervised manner. Although the scope of this work does not include this idea, it might be worthwhile to consider for extended research.

5.6 Reachability analysis using an adjacency matrix

The (i, j) element of an adjacency matrix \mathbf{A}^k gives the number of k -step edge sequences from node i to node j . This property can be used to calculate a reachability matrix \mathbf{R} from an adjacency matrix:

$$\mathbf{R} = (\mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \dots + \mathbf{A}^N)^{\#} \quad (5.7)$$

where the $\#$ refers to the Boolean equivalent of the matrix.

The (i, j) th element of \mathbf{R} indicates that there exists a directed path of any length whatsoever from node i to node j . Jiang et al. (2009) used the reachability matrix of a control loop digraph, indicating only potential direct interactions among control loops, in order to develop a method that studies the propagation of oscillating disturbances through a plant

via control loops. The exponent at which an entry first appears in the reachability matrix indicates the length of paths (degree of separation) between nodes. Maurya et al. (2003) used this idea to reduce the number of spurious results, favouring a short path to longer ones. Path length, as well as the associated time delay determined by cross-correlation, have also been used to detect the most likely propagation path of disturbances (Bauer and Thornhill, 2008).

5.7 Direction of analysis

There are two different directions in which the eigenvector centrality can be used to assess importance. Some applications, such as the ranking of web pages, attribute importance based on the quantity and significance of inbound connections to a specific node, or references back to a particular page. In the context of determining which node has the most considerable extent of causal influence on other nodes, it is desired to attribute importance based on the quantity and significance of outgoing connections from a specific node to the rest of the system.

The extended SDG approach of Vianna and McGreavy (1995) uses back-propagation to ensure that the analysis misses no faults. Maurya et al. (2006) made use of a combination of forward- and backwards-reasoning.

The convention used in defining the directionality of entries in the adjacency or weight matrices according to row and column entries, as well as whether the right or left eigenvector problem is solved determines whether the ranking results refer to the inbound or outbound metric.

In this work, eigenvector centrality always refers to the results obtained by calculating the right eigenvector and the convention to arrange the adjacency matrix with source (causal) nodes along the columns and destination (affected) nodes along the rows as indicated in Equation 5.1 is followed. Under this arrangement it is necessary to use the transposed weighted digraph adjacency matrix \mathbf{M} as the ranking weight matrix \mathbf{W} in 5.3 to get the desired result of ranking based on outgoing connections. Let

$$\mathbf{W} = \mathbf{M}^T \tag{5.8}$$

where \mathbf{M} is the ranking matrix with directionality defined according to Equation 5.1.

5.8 Normalisation

It is important that the outgoing connection edge weights from each node be normalised so that they sum to unity before the calculation of the ranking eigenvector. In other words, matrix ranking weight \mathbf{W} should be column-stochastic.

Normalisation is important to:

- maintain a consistent interpretation of the ranking results,
- make results compatible with theory related to Markov chain transitions on networks, and to
- provide scale with regards to the biasing network.

As the \mathbf{M} matrix is the result of a linear combination of the \mathbf{G} and \mathbf{Q} matrices, the effect of normalisation operations on these matrices needs to be considered. The magnitude of transfer entropy values between time series signals can vary significantly, depending on various properties (see Chapter 6). The PageRank algorithm assigns binary weights to all connections between nodes and ensures the stochastic property by dividing by the number of links involved. Therefore, the weights associated with the incoming connections to a specific node inherently sums to one. However, in this application, the weighted digraph that associates each connection with some weight which does not inherently sum to one. The question now arises whether the weights should be normalised before weighing the network with the reset probability matrix, or only afterwards?

In the case of a source node affecting only a single node, normalising before weighing with the reset probability matrix will result in a loss of scale relative to the other weights. On the other hand, normalising only after weighing the matrix destroys the true nature of the reset probability, which is required to remain intact if there is a desire to superimpose meta-data as discussed in Section 5.5.

To preserve the meaning of reset probability, the gain matrix \mathbf{G} is normalised before weighing it with the reset matrix \mathbf{Q} to obtain the final gain matrix used for calculating the ranking eigenvector \mathbf{M} . See the section on dummy nodes (Section 7.2) for an alternative attempt at addressing the loss of scale issue.

Row normalisation involves dividing each element of a matrix by the sum of the elements in the same row as shown in Equation 5.9.

$$X_{ij_{\text{norm}}} = \left(\frac{X_{ij}}{\sum_i X_{ij}} \right) \quad (5.9)$$

For a simple equally distributed reset matrix, the value of each entry should merely be unity divided by the number of variables in the network (the dimension of the matrix).

However, for a reset matrix with different rankings, the importance of independent normalisation becomes more apparent. The reset matrix \mathbf{Q} should not be column-normalised before being linearly combined with the gain matrix to form the intermediary weight matrix \mathbf{M} - this will destroy the relative node influence information. However, if the reset vector happened to have equal entries, it will happen to be column-normalised as well.

If dangling nodes are present in the network, some rows of \mathbf{G} will sum to zero instead of unity. These rows will not result in a row-normalised weight matrix \mathbf{M} after linear combination with a unity sum distribution of weights with the reset matrix \mathbf{Q} , even if the reset matrix happened to be row-normalised. There is, therefore, need for a final row-normalisation step. After the \mathbf{M} matrix have been transposed to create matrix \mathbf{W} , this row-normalisation is achieved by a column-normalisation operation.

Owing to the challenge of correct normalisation, it is necessary to follow the above scheme accurately in the sense that the transpose operation is only performed once on the combined \mathbf{M} matrix to ensure the desired direction of analysis, instead of doing it separately for the gain and reset matrices.

5.9 Implementation of ranking algorithms

Different algorithms for calculating the desired graph ranking metrics are more efficient depending on the density of the graph. One of the primary purposes of a knowledge-based network generation (if employed) is to reduce the density of the graph, making computation faster, more efficient, and resulting in less spurious results generated from data-based methods.

Power iteration (also referred to as Von Mises iteration) is a standard algorithm for determining the dominant eigenvector. It is efficient on sparse graphs. Although easy to implement, for the sake of robustness the methods available for calculating eigenvector as well as Katz centrality in the NetworkX Python package (Hagberg et al., 2008) is used in the accompanying software.

5.10 Network reduction

It is often the case that the network connectivity is too high to allow for clear visualisation. In these cases, it is helpful to have some means of indicating the main paths amongst variables. Longest path search methods fail on graphs that contain cycles.

The following steps for achieving reduced graphs is applied to obtain clear visualisations:

1. Delete all edges with weights below a specific threshold.

2. Delete all low order edges.

Specifying a threshold as a specific percentile of the weights presents on the graph edges, for example, 75%, is recommended. Choosing the threshold too high will eliminate some essential structural paths.

The degree of separation up to which the algorithm performs the search for higher order connections can be specified. Alternatively, the accompanying software implementation provides the option of searching until no unique children nodes remain uninvestigated.

An option for "weight discretion" is made available, whereby no elimination of a higher order connection between a source and child node occurs if this connection weight is higher than that between the last higher-order child to the destination node under question.

Part II

Exploration

CHAPTER 6

INFLUENCE OF TIME SERIES PROPERTIES

This chapter discusses the effect of varying signal properties typically encountered in process engineering on the behaviour of selected causal interaction metrics.

Evaluating causal links from process data is not trivial. Process dynamics exhibit various varying properties that make it difficult to get consistent results. Some of these features are relatively uncommon in other disciplines. Their impact on the final values obtained together with the required interpretation needed has not been described extensively in the current literature, in particular for some of the more modern and sophisticated measures, such as transfer entropy and related estimation methods.

Chapter 4 mentions some of the difficulties involved in calculating transfer entropy on time series data. In addition to the fundamental challenges encountered, process time series data exhibits some unique features not frequently encountered in other disciplines, such as:

- process time constants vary from a few seconds to several hours,
- time delays of up to several minutes are common, and
- not all measurements are sampled at the same frequency.

One of the simplest test cases possible, namely Gaussian distributed random numbers passed through FOPDT transfer functions with varying time constants, is used as a basis for this initial analysis. A fixed-step ode1 (Euler) solver was used to generate the data. The variables in play include:

- time constant of the transfer functions,
- transport delay,
- sample time of the random number generator,
- variance of the random number generator,
- sub-sampling interval used in the calculation,
- number of samples used to calculate the transfer entropy, and
- simulation time step used to simulate the system.

The simulation time step is fixed at some small value for the results to be sufficiently accurate. Since there is no equivalent varying parameter in the real world, studying the simulation step effects is pointless. The effect of all other parameters on the simple and directional transfer entropy using both kernel and KSG estimators as well as cross-correlation values is investigated.

This study follows an experimental design whereby a base case is defined and each variable of interest varied in turn. The base case parameters are as follows:

Table 6.1: Base case parameters

Variable	Base value
Process time constant	1.0
Gaussian noise generator sample time	0.1
Gaussian noise generator variance	0.1
Simulation time step size	0.01
Sub-sampling interval	1
Sample size	2000
Delays tested	200
Transport delay	1

The test configuration reads data from a starting index corresponding to a time position further than the transport delay to avoid the initial portion of the data where the output of the processes is zero. The total simulation time is irrelevant as long as enough samples are generated to allow for the initial position, the sample size, the sampling interval and the range of time delays investigated. The minimum number of samples needed to run an analysis based on the above parameters is described by

$$T = \frac{S + N + D}{I}, \quad (6.1)$$

where T is the total amount of data points required, S is the starting index, N is the sample size, D is the number of delays investigated (regarding samples in the sub-sampled set), and I is the sampling interval. For example, if the sample size is 2000, sub-sampled

at an interval of 5 samples starting from index 1000 and it is desired to investigate 40 delays (in terms of samples in the sub-sampled set), then the total amount of data points required is $(1000 + 2000 + 40) \times 5 = 15\,200$.

Interaction measures investigated include cross-correlation and the simple and directional transfer entropies calculated according to the kernel and KSG estimators, with and without auto-embedding to condition on self-information in the latter case.

In the following tests all signals were standardised.

6.1 Time constant

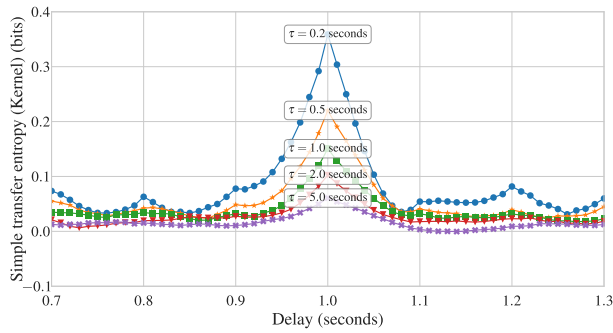
Figure 6.1 presents interaction measures over a range of delays for process time constants $\tau = [0.2, 0.5, 1.0, 2.0, 5.0]$.

In cases (a) to (d), a clear peak is visible in the transfer entropy measurements at the exact real transport delay of one second, indicating that maximising TE without auto-embedding is useful for accurate estimation of the transport delay between highly active signals.

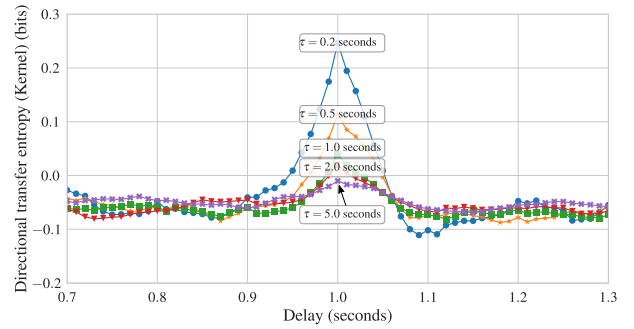
This result indicates that due to the extreme sensitivity of transfer entropy to transport delays, it is imperative to maximise the causality measure over a suitable range of delays to get a representative value to use as a measure of causal relationship strength between two elements. However, rankings that optimise for time delays are sensitive to the search range allowed. If the maximum delay is too short, there is a risk of missing meaningful relationships, whereas if it is too long, the method might indicate affected variables as causes for the variables that affected their behaviour in the first place.

Directionality tests are designed to reduce the occurrence of effects passing as sources. Note that the directional transfer entropy measure returns negative values if the estimate of transport delay is too long because the actual source node now lags behind the affected node. However, also note that the maximum negative value is significantly smaller than the maximum positive value, correctly indicating that the signal contains less information after being passed through the first order transfer function, and does not predict the noise as well as that the noise predicts the filtered signal.

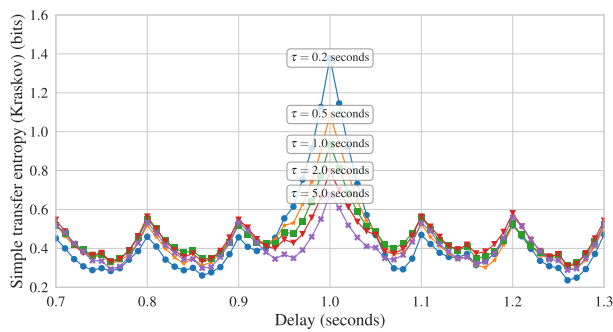
Cross-correlation is also sensitive to transport delays but fails to estimate the actual delay precisely, with the maximum value position depending on the process time constant. The indicated delay is closer to the real delay for shorter time constants. This result agrees with the observations recorded by Tabaru et al. (2005) who showed that process dynamics, in addition to time delays, contribute a shift in the peak of cross-correlation values between input and output variables. The peak of the cross-correlation function has an additional delay θ equal to the magnitude of the slope of the phase graph of the



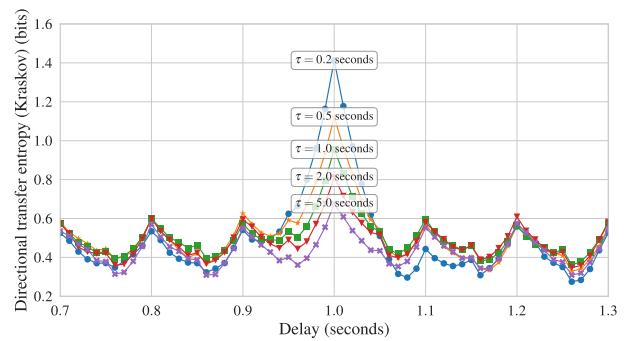
(a) simple; TE kernel estimator



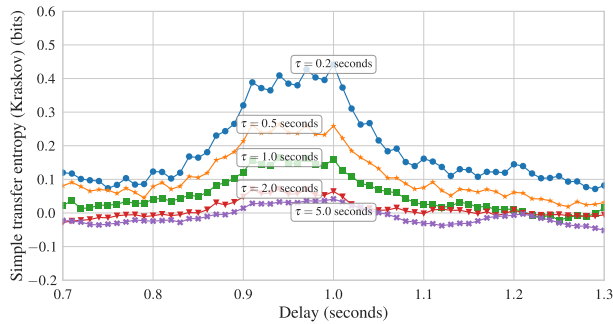
(b) directional; TE kernel estimator



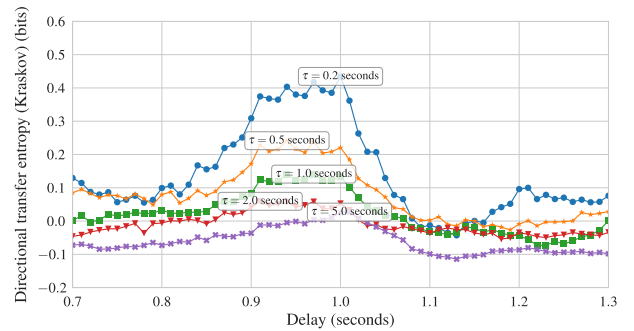
(c) simple; TE KSG estimator



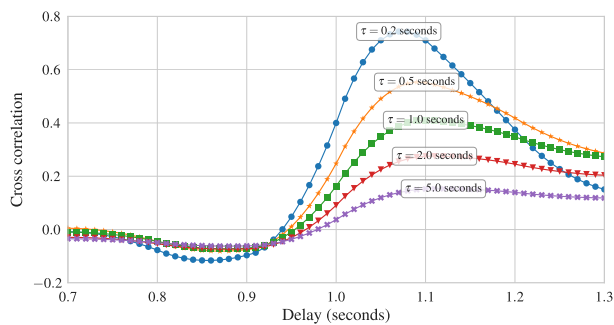
(d) directional; TE KSG estimator



(e) simple; TE KSG estimator; autoembedded



(f) directional; TE KSG estimator; autoembedded



(g) cross correlation

Figure 6.1: Interaction weights over a range of time delays for different process time constants [standardised]

process at the dominant frequency of the disturbance. For signals resembling random noise, transfer entropy is a more useful measure than cross-correlation.

For both the simple and directional transfer entropies a clear relationship between the process time constant and the maximum value of the calculated transfer entropy is observed. This relationship is very well approximated by a power relationship, especially in the case of simple entropies, and therefore results in a straight line on a log-log plot.

If it is desirable to eliminate the effect of process time constant on the value of the causality measures, the researcher might consider attempting a possible measure adjustment involving dividing the calculated value by the time constant raised to the power of the gradient of the fit on the log-log plot. As always, it is important to note that the results presented above will only be valid for signals which are similar to Gaussian noise. The causality measures can potentially return very different results for signals that contain more structure.

Whether it is indeed desirable to adjust the causal measure for the effects of process time constant will depend on the situation. On the one hand, adjusting for the process time constant will diminish the attenuation of causal paths in the network due to the presence of large process time constants in individual sections. To do this an estimate of the dominating time constant of the process will be required, which can introduce error into the method. On the other hand, fault-carrying signals that pass through processes with long time constants will have fewer high-frequency elements. It is easier to reject low-frequency disturbances compared to high-frequency ones and these are, therefore, likely to be less of a concern than those that are not filtered by a considerable time constant process. If the analyst desires that the ranking should reflect this effect, no adjustment of the causal measures should be made to compensate for the influence of different process time constants. However, the user will do well to take note of this effect and keep it in mind as a possible explanation for weak causal values between elements that are suspected to be highly causally related.

6.1.1 Kernel vs. KSG estimation

The magnitudes of the kernel and KSG estimates differ significantly. The fact that the kernel estimator is sensitive to the bandwidth selected (0.25 in all these tests) explains this effect. Note that in the case of directional transfer entropy calculated according to the kernel estimator, the value never becomes positive for large process time constants, while the KSG estimator remains positive at all delays. Even though the kernel estimator value becomes negative according to the directional measure, the significance threshold (not shown) also becomes negative, such that measure passes the significance test.

The KSG estimator results are better behaved. The KSG estimator results exceed their

significance threshold (which always remain positive) by a significant margin compared to the kernel estimator results. Interestingly, the KSG directional results are very similar to that of the simple results, indicating that this estimator correctly assigns a near-zero value to the inverse relationship. Although the kernel measure shows sub-harmonic peaks, the KSG estimator shows them at a higher frequency due to the adaptive resolution of the estimator.

6.1.2 Effect of auto-embedding

Enabling auto-embedding significantly reduces the absolute magnitude of results obtained, removes the subharmonic peaks and gives a plateau of similar results before the real delay instead of a well-defined peak. Not conditioning on self-information in the signals will result in over-estimation of the transfer entropy between them; therefore this decreased magnitude is expected. For the cases of $\tau = [0.5, 1.0, 2.0]$, the plateau effect is so strong that delays other than the real delay maximise the measure by a narrow margin.

If auto-embedding is enabled, the validity of the directional transfer entropy measure is dubious as the search optimised the parameters for the source and target signals according to different considerations. The directional measurement does not differ significantly for the KSG estimator, and since it might be even misleading in this case, it is recommended that only the absolute measurement should be used for the KSG estimator, especially when conditioning on self-information.

It is interesting to note that for all cases except $\tau = 5.0$, the same embedding dimension and time step was optimised for, while the search reached the pre-configured limit of 5 in the last case, which also had a negative significance threshold. It is essential to study results and check whether the limits used were relevant to the relative dynamics of the processes investigated. However, increasing the optimising search space increases computational expense.

Further investigation into allowing for a broader search space revealed concerning behaviour, with the transfer entropy measurements falling close to zero with high auto-embedding dimensions. Whether this will occur for other signals apart from Gaussian noise remains to be investigated. It is possible that alternative self-information conditioning schemes such as AIS will be better behaved.

6.1.3 Effect of relative time captured by sample

Note that in the examples above, the total time span covered by the time series data vector (20 seconds) was much larger than the process time constants (maximum of 5 seconds).

Figure 6.2 demonstrates the effect of extensive time constants such that the time span of data used in the estimate only covers a fraction of the process time constant and therefore only captures effects from the initial dynamics.

The causality measures seem to tend towards a lower limit, with no significant difference between results observed even after the time span covered is smaller than the process time constant.

Weber et al. (2017) recently published a rigorous study on the effects of filtering and down-sampling on the estimation of transfer entropy.

6.2 Noise generator sampling interval

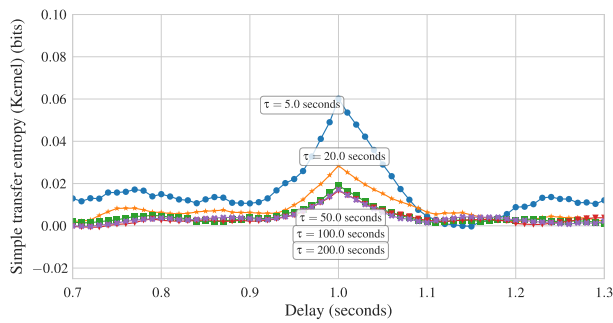
The sampling interval of the noise generator has a significant effect on the values obtained from both transfer entropy and cross-correlation. According to the Nyquist sampling theorem, a signal needs to be sampled at a minimum of half the period corresponding to the fastest frequency present in the signal to ensure that no aliasing effects come into play.

Figure 6.3 displays the response of investigated measures to different noise generation sampling intervals over a range of time delays. The signals passed through a process with a time constant of one second. For the kernel and KSG estimators without auto-embedding, shorter noise sampling intervals return larger transfer entropy values and sensitivity to the correct sample delay reduces with increased noise sampling rate. Note that for an interval of 0.01 seconds, in which case the simulation time step was short of the Nyquist frequency for proper sampling of this signal, only the sample corresponding to the actual delay indicated any significant transfer entropy. The auto-embedded results failed to detect a unique optimal delay for sampling rates lower than the Nyquist frequency.

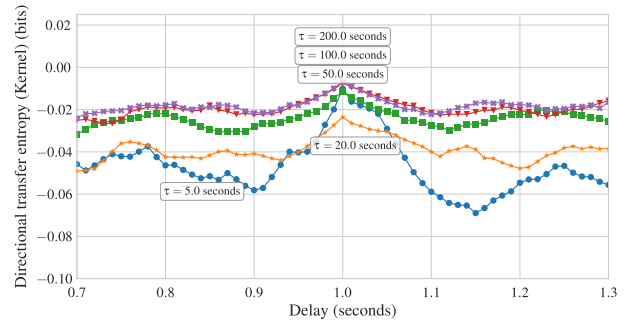
As far as the magnitude of the values is concerned, cross-correlation shows a different response to the noise sample interval as transfer entropy, returning smaller values for smaller sampling intervals. However, the actual delay can be detected much better from the results corresponding to shorter sampling intervals.

6.3 Noise generator variance

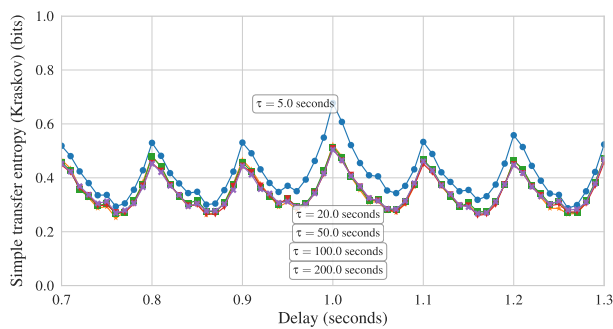
It comes as no surprise that, if the data is normalised, noise signals with different variances do not affect the returned values. This effect brings into question the wisdom of normalising data, as one would hope in the context of identifying the most important causal factors that the method will assign stronger weights to signals that show more variance compared to those that show less, everything else being constant.



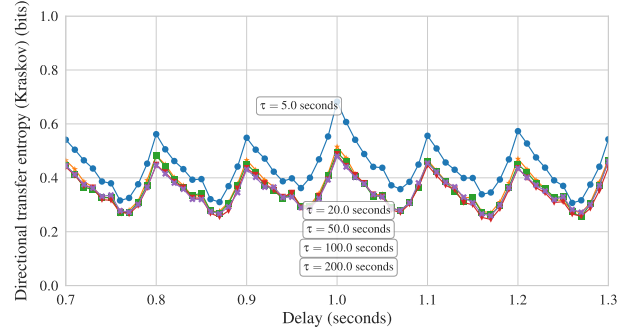
(a) simple; TE kernel estimator



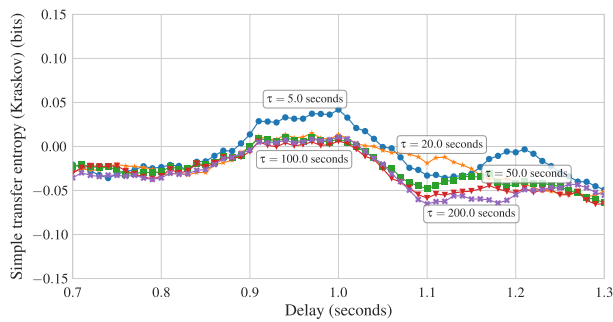
(b) directional; TE kernel estimator



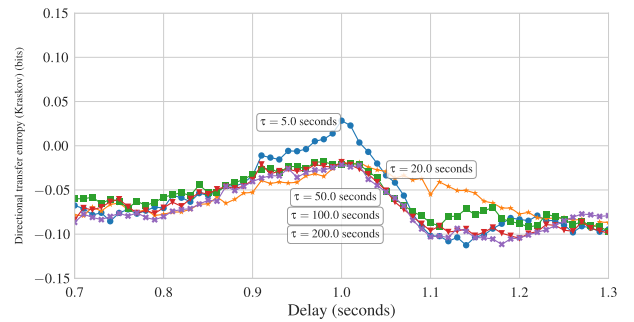
(c) simple; TE KSG estimator



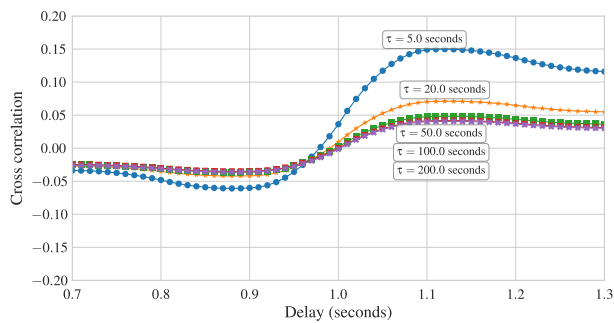
(d) directional; TE KSG estimator



(e) simple; TE KSG estimator; autoembedded

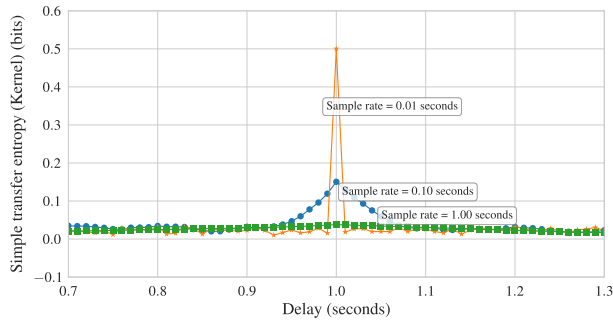


(f) directional; TE KSG estimator; autoembedded

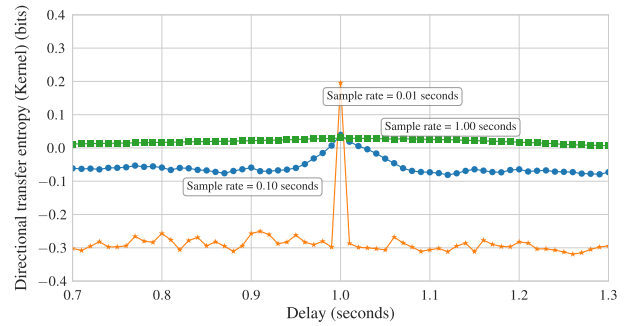


(g) cross correlation

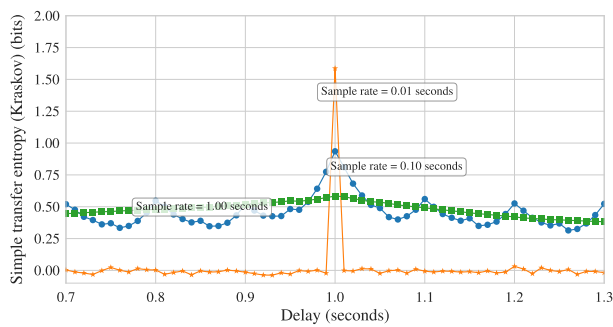
Figure 6.2: Interaction weights over a range of time delays for different process time constants that far exceed the time span covered by analysis [standardised]



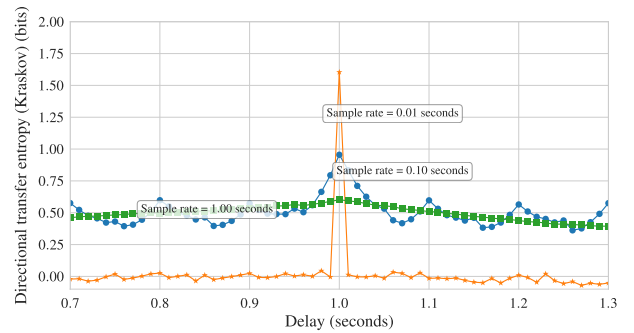
(a) simple; TE kernel estimator



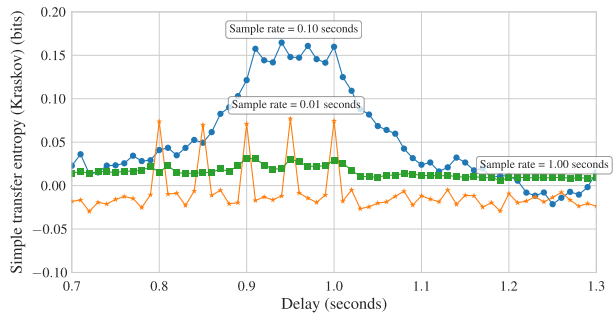
(b) directional; TE kernel estimator



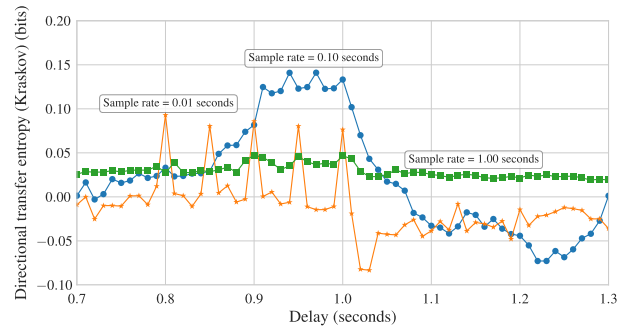
(c) simple; TE KSG estimator



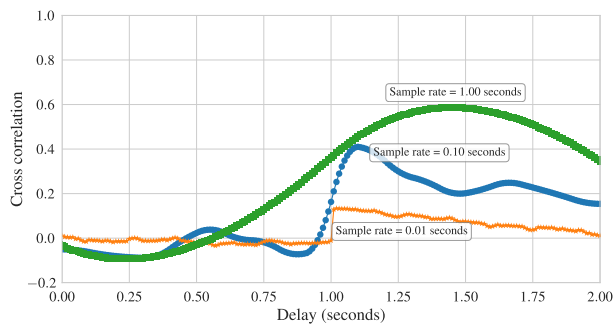
(d) directional; TE KSG estimator



(e) simple; TE KSG estimator; autoembedded



(f) directional; TE KSG estimator; autoembedded



(g) cross correlation

Figure 6.3: Interaction weights over a range of time delays for different noise sampling rates [standardised]

Not normalising the data before calculations lead to odd results, returning values differing several orders of magnitude for different signals. Therefore, some form of normalisation is needed. Section 4.8 discusses a potential workaround for the normalisation problem.

The effect of normalisation is more likely to result in false positive significance indicators as opposed to false negatives. For the time being, the analyst is recommended to manually inspect the variance of signals if a false positive is suspected (see Section 9.5.2).

6.4 Sample size

Figure 6.4 presents results of sample size on values calculated using the various measures employed. The simulation time interval was kept constant; therefore longer samples were exposed to more original data than the shorter ones.

The first significant result to take note of is the fact that being exposed to more data with the same properties had minimal effect on the values of the results returned for the kernel estimator. Improved stability is the primary effect associated with increasing sample size. Note that small samples result in the indication of effects at harmonic periods of the actual delay.

The KSG estimator showed different behaviour, remaining more stable at smaller sample sizes and converging after 1000 samples, except in the case of auto-embedding. There is a positive correlation between transfer entropy at the optimising delay and sample size.

No difference occurs for sample sizes as small as 200 if the source signal is a pure sinusoid. The fact that a single oscillation at sufficient resolution full represents the PDFs involved explains this effect. Adding more samples does not change the PDF estimates, and therefore adds no contribution to the analysis.

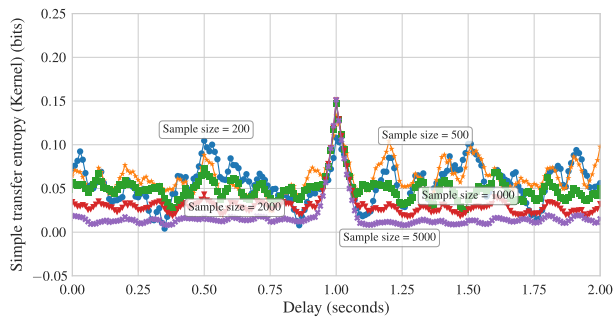
The correlation measure is not significantly affected by sample size.

6.5 Sampling rate

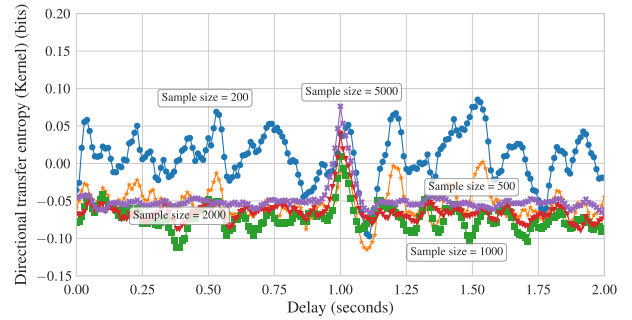
There is a complex interaction between the effect of noise generator interval (the dynamics of the source process), the data sampling interval and time constant of the FOPDT filter.

Figure 6.5 shows how the magnitude of transfer entropy at the correct delay increases with lower sampling frequencies for all estimators, while correlation values are relatively unaffected.

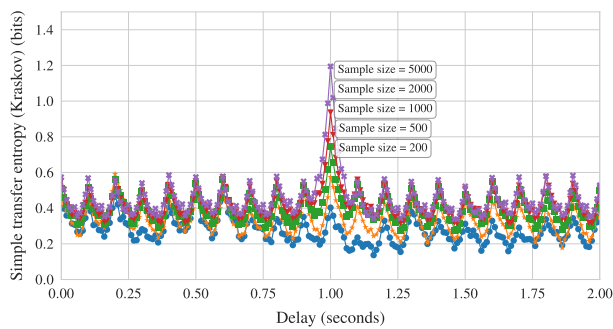
Figure 6.6 shows a very significant interaction between the maximum causality estimate values measured at the true delay for different sampling intervals over a range of process time constants.



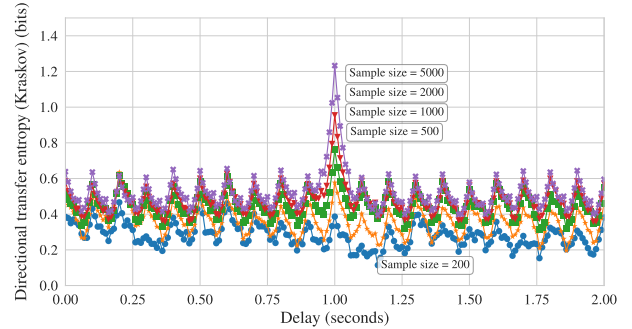
(a) simple; TE kernel estimator



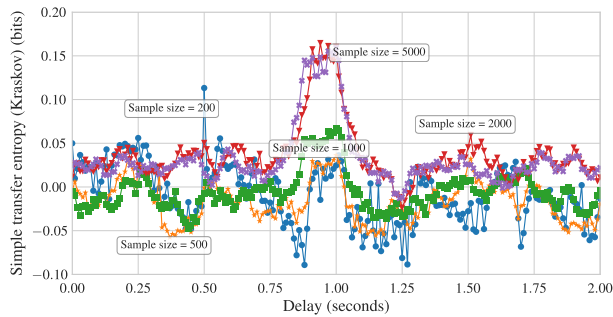
(b) directional; TE kernel estimator



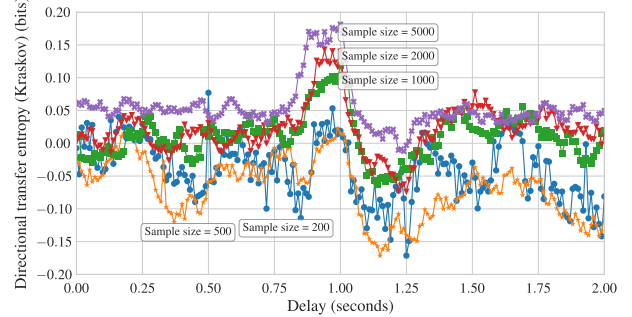
(c) simple; TE KSG estimator



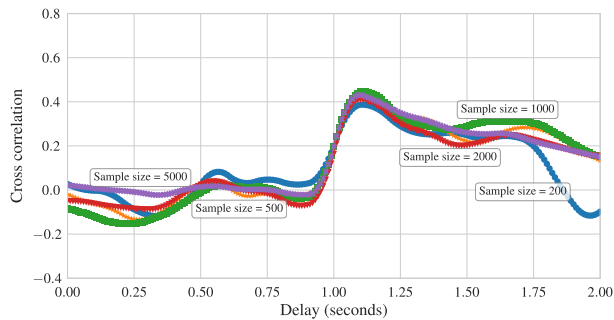
(d) directional; TE KSG estimator



(e) simple; TE KSG estimator; autoembedded

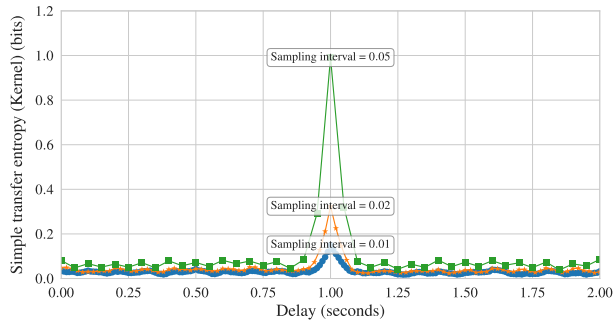


(f) directional; TE KSG estimator; autoembedded

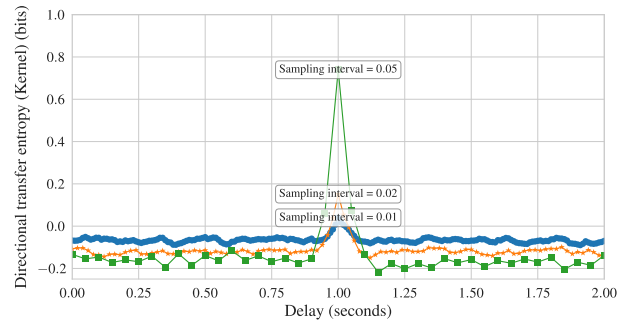


(g) cross correlation

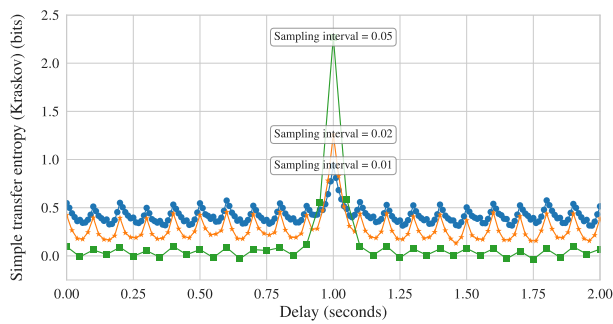
Figure 6.4: Interaction weights over a range of time delays for different sample sizes [standardised]



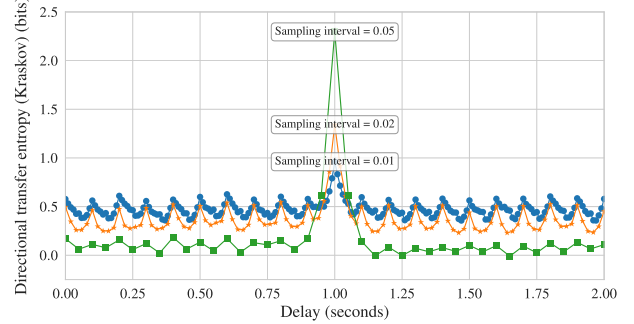
(a) simple; TE kernel estimator



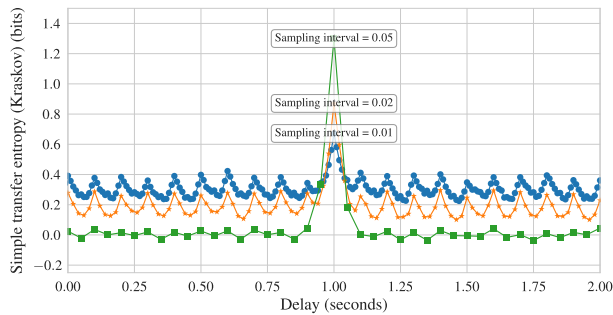
(b) directional; TE kernel estimator



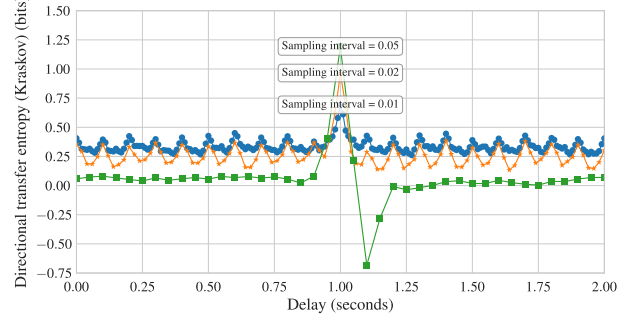
(c) simple; TE KSG estimator



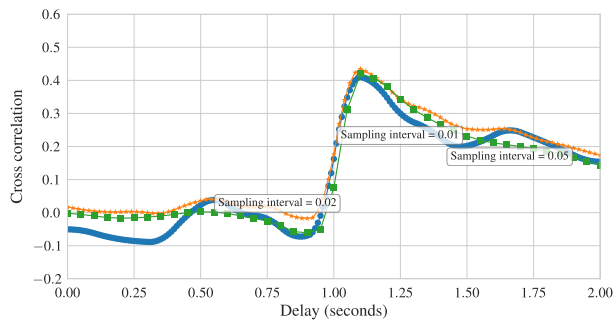
(d) directional; TE KSG estimator



(e) simple; TE KSG estimator; autoembedded

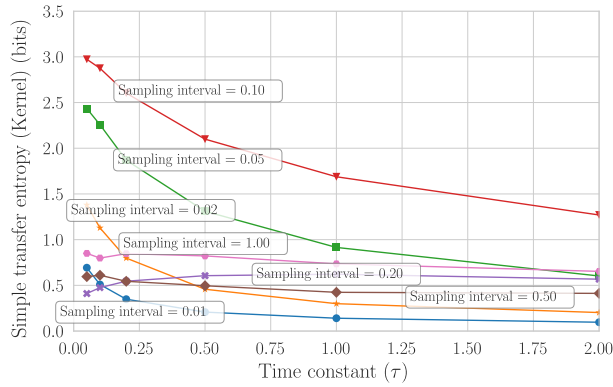


(f) directional; TE KSG estimator; autoembedded

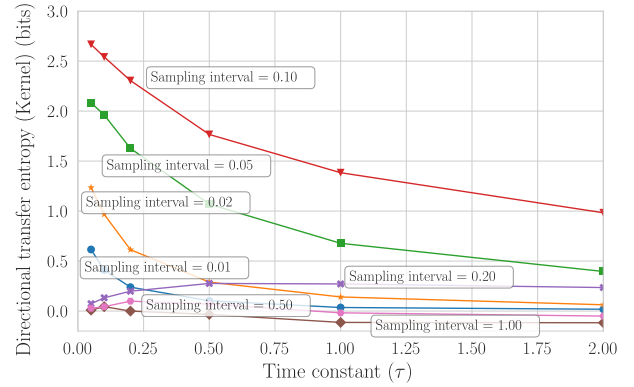


(g) cross correlation

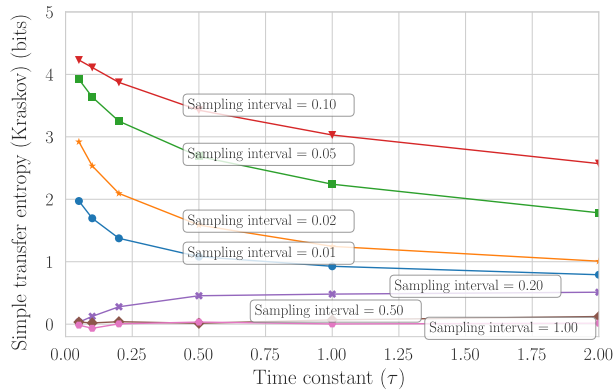
Figure 6.5: Interaction weights over a range of time delays for different sampling rates [standardised]



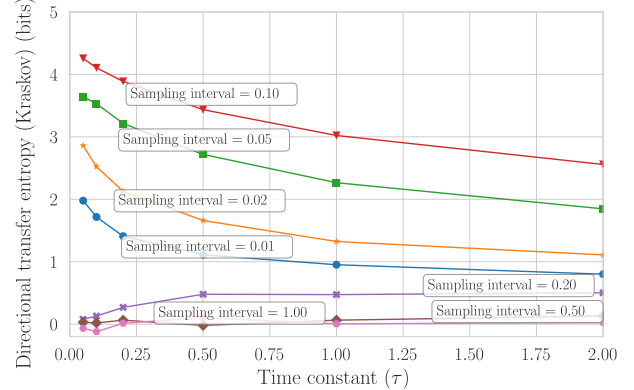
(a) simple; TE kernel estimator



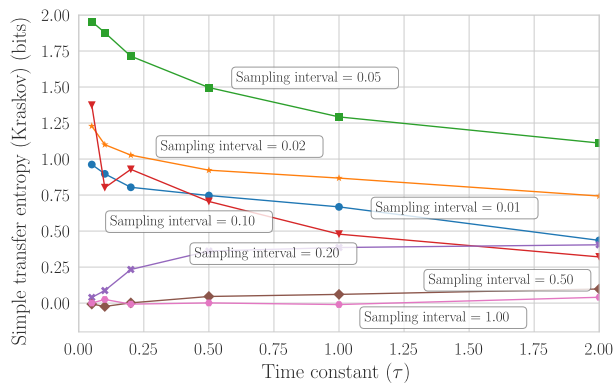
(b) directional; TE kernel estimator



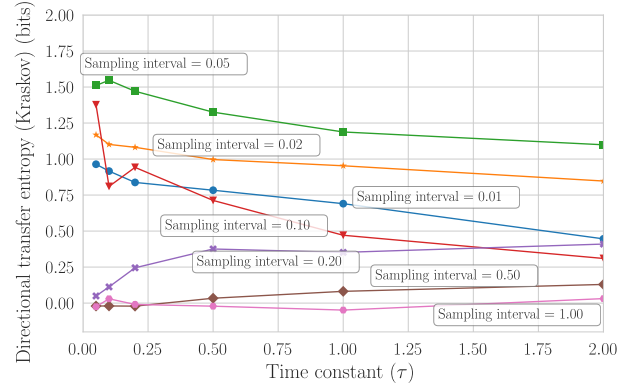
(c) simple; TE KSG estimator



(d) directional; TE KSG estimator



(e) simple; TE KSG estimator; autoembedded



(f) directional; TE KSG estimator; autoembedded

Figure 6.6: Interaction weights over a range of time constants for different sampling rates [standardised]

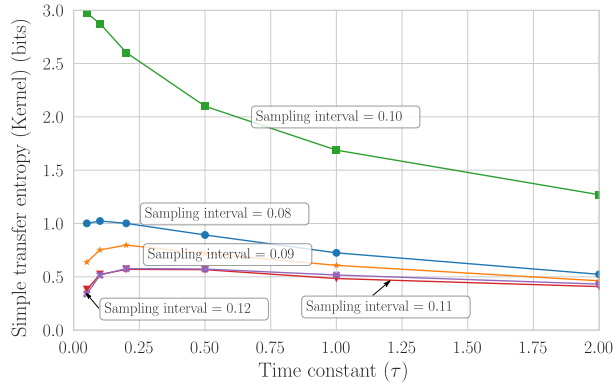
Where the sampling interval is shorter than dictated by the Nyquist sampling theorem with respect to the noise source, longer sampling intervals result in larger transfer entropy measures. As with the analysis of different simulation time steps the sample sizes remain constant, exposing data sampled at different sampling rates to proportionally varying time periods. The fact that slower sampling intervals result in higher transfer entropy rates can likely be explained by the fact that the overall entropy contained in these datasets are higher, having been exposed to more random samples.

A second observation of note is that transfer entropy decreases with increasing time constant. The fact that larger time constants will result in filtering a more significant portion of the higher frequencies out, destroying some information in the process, explains this effect.

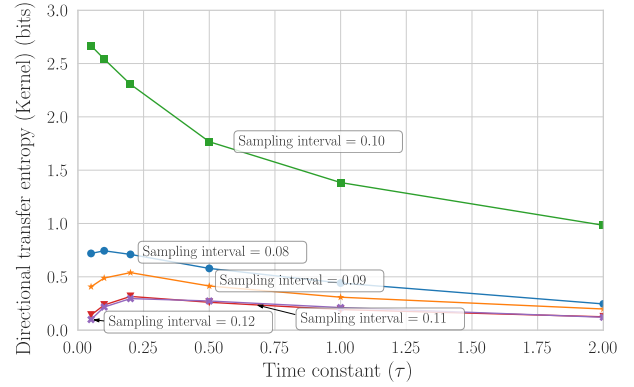
In this figure an inverse relationship is exhibited for a sampling interval of 0.20. Instead of starting at the maximum value possible, the transfer entropy starts at a value close to zero. As the time constant becomes longer, the causality measure initially increases up to a point before starting to decrease again. The trend of initial smaller values only disappears when the sampling frequency is higher than the Nyquist frequency.

To investigate these effects in more detail, Figure 6.7 provides a higher resolution plot of sampling intervals closer to the noise sampling rate. The causality estimates occurring in the region just below the noise sampling rate follow a similar trend to that just after it, with an apparent unique effect occurring when the sampling frequency coincides precisely with the noise sampling rate.

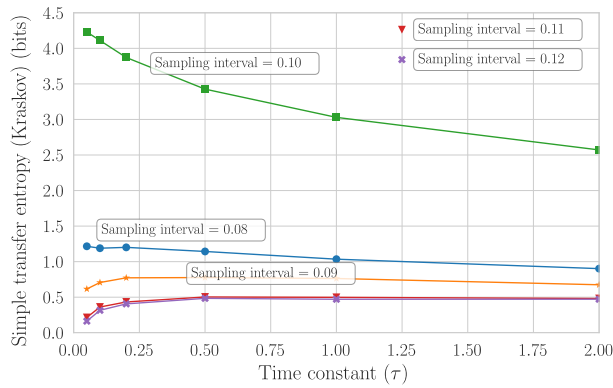
Unsurprisingly, care must be taken to select a sampling interval shorter than dictated by the Nyquist sampling theorem in the context of signals of interest. However, as there is also a desire to optimise the exposure of the measure to as much activity as possible without increasing computational burden with large sample size, choosing a sampling rate much faster than the Nyquist limit is expected to be of little benefit. If the interest is primarily in signals passing through slow processes with long time constants, fast frequencies will be filtered out and longer sampling intervals will not result in loss of fidelity.



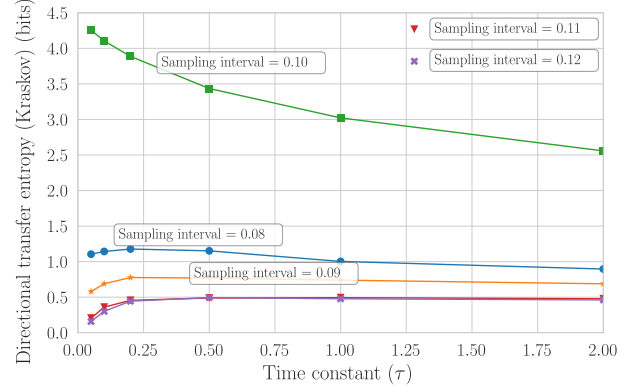
(a) simple; TE kernel estimator



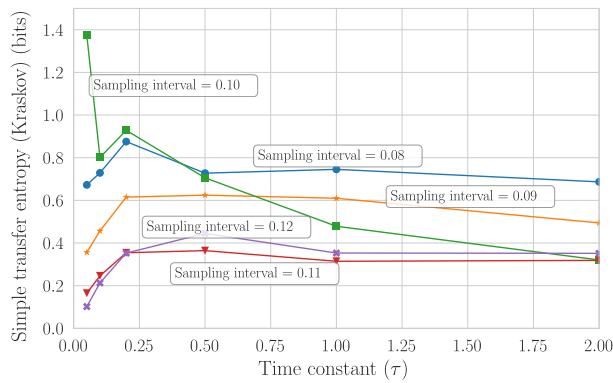
(b) directional; TE kernel estimator



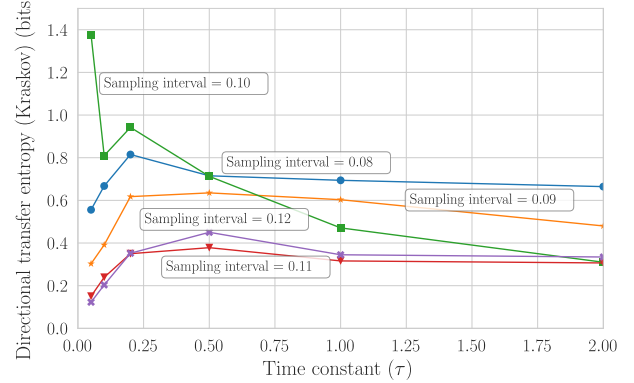
(c) simple; TE KSG estimator



(d) directional; TE KSG estimator



(e) simple; TE KSG estimator; autoembedded



(f) directional; TE KSG estimator; autoembedded

Figure 6.7: Interaction weights over a range of time constants for different sampling rates close to the noise generation interval [standardised]

CHAPTER 7

RANKING METHODS AND PARAMETERS

This chapter explores the behaviour of selected algorithms for calculating relative importance scores of network nodes.

7.1 Incorporating indirect connections

Intuitively it is expected that knowing the real connectivity of the plant and using a sparse weight matrix limiting connections to occur only between nodes that have direct physical or logical connection should give a more definite, distinct ranking result that indicates control related faults with less ambiguity.

However, if the eigenvector centrality measure is applied to calculate the ranking, using a fully connected adjacency matrix (incorporating all weights that pass the significance test), not making use of any structural knowledge of the plant, often delivered better results.

In a fully connected gain matrix where connections between all process elements (graph nodes) are allowed, some higher degree interactions pass the significance test, resulting in the inclusion of indirect connections between the source and target as direct connections in the calculation. Although the chain length of nodes subsequently affected by an affected node does play a very significant and direct role in boosting a source node's importance score as shown in Section 7.3, additional direct connections provide a boost to the source node's importance score and therefore assist in indicating the correct source node.

The advantages of integrating known plant topographical information, especially with regards to reducing the occurrence of spurious results and allowing for clear propagation paths useful for troubleshooting to be indicated, motivated the search for a potential workaround that would allow sparse matrices to give more distinct results.

Section 5.3.5 introduced the use of an alternative ranking algorithm referred to as Katz centrality that ascribes value to higher- as well as first-order connections, boosting the importance of nodes that is close to the start of a long pathway. Section 5.6 introduced the concept of the reachability matrix and use of the adjacency matrix's exponent to count the number of connections between nodes with higher degrees of separation. The exponent of a weighted gain matrix is the sum of products of weights along the different paths up to the order calculated. Katz centrality makes use of this fact to incorporate the importance of higher order connections in determining a centrality score for a node.

Increasing the attenuation factor has the two-fold effect of 1) decreasing the default starting importance for nodes and 2) increasing the gain in importance associated with traversing an edge in a chain. The practical implications of these statements are clarified by example in Section 7.3.3.

The Katz centrality will only provide more distinct rankings than eigenvector centrality if the attenuation factor is larger than unity. As the largest eigenvector of a column-stochastic matrix is always identically unity, it only makes sense to apply Katz centrality to the original gain matrix, and only if it does contain dangling nodes - which can almost always be expected to be the case for a real plant section terminating at some battery limit. The eigenvalues for the disconnected gain matrix in the case under consideration are all zero, and therefore the Katz attenuation factor can be made as large as desired. The starting importance of nodes tends to zero in the limit that the attenuation factor tends to the reciprocal of the maximum eigenvalue of the weight matrix from below. In the limit as the Katz attenuation factor approaches infinity, all the importance will be associated with the source node only.

As incorporating a reset probability matrix will always force the maximum eigenvalue to a value of unity, there is no well-defined way of incorporating additional information via a reset bias in the current implementation of using Katz centrality as is the case for eigenvalue centrality.

7.2 Dummy nodes

The potential concern about loss of scale due to normalisation for nodes with a single outward connection is discussed in Section 5.8. Wilken (2012) suggested one potential solution to this problem by including “dummy nodes” in the weighted digraph at all

locations where a node has a single outgoing connection. The inclusion of this dummy weight will place the other single connections in perspective relative to each other, based on their absolute values. The calculation of the final relative importance scores disregards the dummy nodes.

The dummy node edge weights should be an appropriate value that will allow for a useful comparison between the weights on all edges associated with nodes. If the measures used to determine the edge weight as an absolute maximum with a specific meaning, such as is the case for cross-correlation, a unit weight may be assigned, for example. The ratio between unity minus the damping factor, $1 - m$, and the dummy weight was found to be the most important relation in determining the strength of the dummy weight effect. Setting the dummy weight to a tenth of the inverse of difference between unity and the dampening factor returned reasonable results. For example, if $m = 0.99$, then the dummy weight should be set to $DW = 10$. The sensitivity improves asymptotically in the limit that m approaches unity and DW approaches infinity.

Attempting to compensate for the loss of scale by use of the proposed dummy node scheme is questionable if the weight measure values are strongly affected by various properties of time signals represented in a broad spectrum in the data set, such as is the case for the information-theoretic causality estimation measures. The use of dummy nodes should remain limited to evaluations involving cross-correlation in the scope of the analysis made possible by the implementation described in this work. Example applications of dummy nodes are presented in Section 7.3.4.

7.3 Sample rankings

The following section presents some simple examples to help the reader develop an intuition for how the node ranking scores are affected by network structure and available tuning parameters.

7.3.1 Eigenvector centrality

A fully connected weight matrix with unit weights in all entries is arguably the most trivial test case possible. Naturally, each node is expected to have equal importance.

A slightly more interesting test case is that of nodes connected in a series. The importance should increase linearly, with the ranking assigning the highest importance score to the node at the very start of the chain of influence. As a more substantial damping factor will spread the node ranking importance, this demonstration uses the limit as the damping factor approaches unity. The resulting normalised rank array is presented in Table 7.1.

Table 7.1: Eigenvector centrality node importance distribution for chain

Variable	Importance score
X_1	$5/15$
X_2	$4/15$
X_3	$3/15$
X_4	$2/15$
X_5	$1/15$

As can be observed, this ranking follows the trend where each node has an equally distributed starting importance, and the importance of each node is the sum of the importance of all nodes it points towards plus its initial importance.

7.3.2 Biased eigenvector centrality

To investigate the effect of biasing, consider the case of a fully connected 5×5 system, with the third node being assigned a relative reset bias weight of 3, whereas the other nodes are assigned a value of unity. If the dampening factor is maintained at the limit as it approaches unity, the reset bias vector importance does not affect the ranking results. If the dampening factor is set to zero, the third node will have a normalised importance $3/7$ precisely three times as big as the other nodes $1/7$.

If the system contains dangling nodes, the reset bias vector can have an influence even when the dampening factor is set to the limit as it approaches unity. Table 7.2 contains normalised rank array results for five nodes connected in series, a dampening factor of 0.999 and a reset bias vector which assigns a relative importance that is three times greater to the third node.

Table 7.2: Eigenvector centrality node importance distribution for chain with bias

Variable	Importance score
X_1	$7/21$
X_2	$6/21$
X_3	$5/21$
X_4	$2/21$
X_5	$1/21$

Notice that the same patterns observed for the case of the unbiased series apply, namely that the importance of each node is the sum of the importance of all nodes it points

towards plus its initial importance, except that in this case, the third node had a starting importance three times greater than the rest.

7.3.3 Katz centrality

As seen in Section 7.3.1, for the case of a simple chain the importance scores increase linearly when using the eigenvector centrality measure together with an equally distributed reset matrix. However, as nodes at the source of a long path of interaction are likely to be of vital importance, it is desirable to favour them to ensure rankings are as distinct as possible.

The Katz ranking effect is demonstrated by again considering five nodes connected in series. Recall that no reset probability matrix is related to the current implementation of the Katz centrality method. When using an attenuation factor of $\alpha = 1$ the results obtained is the same as that shown in Table 7.1, demonstrating that Katz centrality can be used to compute the effective (unbiased) eigenvector without creating an entirely connected matrix first. Since the maximum eigenvalue of this rank-deficient matrix is zero, the attenuation factor can be set as high as desired. The results for an attenuation factor of $\alpha = 3$ is shown in Table 7.3.

Table 7.3: Katz centrality node importance distribution for chain with $\alpha = 3$

Variable	Importance score
X_1	$121/179$
X_2	$40/179$
X_3	$13/179$
X_4	$4/179$
X_5	$1/179$

Note that in this case, the importance of a node is equal to three times the sum of incoming importances plus the initial importance equally distributed among all nodes.

7.3.4 Dummy nodes

The simple gain matrix in Table 7.4 demonstrates the potential usefulness of dummy nodes to bring scale to edge weights associated with nodes that only have one outgoing connection.

Table 7.4: Dummy node demonstration gain matrix (no dummies)

	X_1	X_2	X_3	X_4	X_5
X_1	0	0	0	0	0
X_2	1	0	0	0	0
X_3	0	2	0	0	0
X_4	0	0	1	0	0
X_5	0	0	0	1	0

Owing to the way normalisation is performed, the edge weights in a chain structure are irrelevant in determining the node importance scores. As can be observed, there is a stronger connection between node X_2 and X_3 compared to all the other connections. If this occurred while using a measure where the absolute values of the edge weights carry some definite meaning, as discussed in Section 7.2, it might be useful to boost the importance attributed to X_2 due to this strong connection.

The method adds dummy nodes in parallel to all nodes that have a single outgoing edge. Notice that dummy nodes are not attached to dangling nodes. Table 7.5 presents the gain matrix with dummy nodes (with truncated matrix columns as dummies do not connect to any destination nodes).

For this demonstration, an edge weight of $DW = 10$ is assigned to the dummy connections, while the dampening factor is set to $m = 0.99$, following the recommendation mentioned in Section 7.2.

Table 7.5: Dummy node demonstration gain matrix (with dummies)

	X_1	X_2	X_3	X_4	X_5	D_1	D_2	...
X_1	0	0	0	0	0	0	0	...
X_2	1	0	0	0	0	0	0	...
X_3	0	2	0	0	0	0	0	...
X_4	0	0	1	0	0	0	0	...
X_5	0	0	0	1	0	0	0	...
D_1	10	0	0	0	0	0	0	...
D_2	0	10	0	0	0	0	0	...
D_3	0	0	10	0	0	0	0	...
D_4	0	0	0	10	0	0	0	...

Table 7.6 compares the ranking scores for this matrix to that of an equally weighted chain (also analysed according to the dummy node scheme).

Table 7.6: Eigenvector node importance distribution for chain with dummies

Variable	Importance scores		Difference
	2-3 weight: 2	2-3 weight: 1	
X_1	0.3415	0.3393	+0.0022
X_2	0.2811	0.2770	+0.0041
X_3	0.2043	0.2077	-0.0034
X_4	0.1286	0.1308	-0.0022
X_5	0.0445	0.0453	-0.0007

The use of dummy nodes on the case of an equally weighted chain causes importance of the nodes to increase roughly by twice the initial starting value with each additional increase in chain length as opposed to the more linear form observed for the case without dummy nodes. Notice that this is also different from the Katz attenuation effect, which was multiplicative and not additive. A second important observation is that the use of dummy nodes did indeed allow the different edge weight between nodes 2 and 3 to affect the relative importance scores, shifting the values in favour of the nodes occurring behind the heavier weight. Also notice that regular pattern of weight shifts according to the following ratios: +3, +6, -5, -3, -1 and that percentage-wise the importance of nodes downstream from the increased edge have been affected to the same extent.

However, as can be seen, the differences are minimal and are arguably not be useful enough to justify the loss of linearity in the rankings. No dummy nodes are included in the analysis presented in the rest of this work but the discussion might be of interest to potential extended research in future.

CHAPTER 8

EXPERIMENTAL ANALYSES

This chapter presents results of a systematic study performed on a simple simulated process, exploring the relative efficiency and potential pitfalls of the many methods and configurable parameters studied.

As the behaviour of the measures studied in this work cannot be derived analytically, a structured analysis with regards to accuracy, stability and robustness on basic representative processes is necessary. Ideally, this research should assist in developing an understanding of the methods such that a general recommended procedure can be derived that finds balance between configuration and maintenance effort, required accuracy and computational cost.

8.1 Experimental design

There are 16 different possible configurations for performing transfer entropy calculation according to the KSG estimator due to the options shown in Table 8.1. The kernel and correlation methods lack embedding options, and the correlation measure has no direction, such that these methods provide 8 and 4 different possible configurations respectively.

Table 8.1: Weight analysis configuration options

Concept	Options
Scaling	standardised; Skogestad
Significance tested	True; False
Bi-directional delay analysis	False
Directionality entropy	True; False
Auto-embedding enabled	True; False

The examples provided demonstrate the results of these different configurations. Due to the exponential increase in possible experiments with additional variables, other parameters such as sampling frequency and size relating to the nature of time series signals involved as discussed in Section 6 are not varied but chosen to suit the dynamics of the problem. Similarly, a simple and universal node ranking and graph reduction approach is used; see Section 7 for a discussion of other options.

The fixed configuration parameters used in this study is provided in Table 8.2. Example configuration files are included in Appendix C.

Table 8.2: Analysis configuration

Option	Value
Significance test	rank order iAAFT
Delay range	100 (0.5 hour)
Start index	100
Sample size	1800
Sampling rate	0.005 hours
Multiple time region bins	20
Time region size	10 hours (2000 datapoints)
Ranking	eigenvector ($m = 0.85$)
Reduction rules	percentile = 0.50; depth = 5; no weight discretion

8.2 Simulated level and composition control problem

Consider the simple mixing process presented in Figure 8.1. Process parameters, control limits, loop tuning is presented in Appendix A.

This process exhibits many aspects of interest to process engineering, including:

Nonlinearity

The relationship between liquid level h , valve opening v and outflow rate F_C is non-linear.

Interacting control

The composition x and level h control loops are interactive.

Integrating processes

The tank is an integrator of volume, and its level h will have in integrating response to any of the flow rates in isolation.

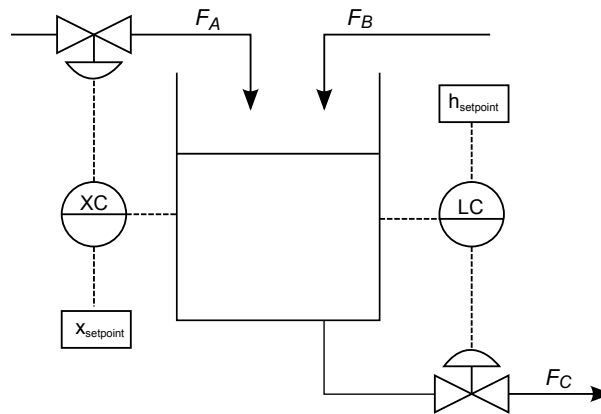


Figure 8.1: Level and composition problem process flow diagram

Table 8.3 presents relevant process variables together with their classification as either an CV, Manipulated Variable (MV), Disturbance Variable (DV) or Process Variable (PV).

Table 8.3: Level and composition process variables

Symbol	Type	Description
F_A	MV	Inlet flow of component A
F_B	DV	Inlet flow of component B
F_C	PV	Outlet flow of product C
h	CV	Liquid level
v	MV	Valve opening
x	CV	Composition

8.3 Disturbances

In order to assess the efficiency of the different methods for detecting the source of a disturbance, two random noise disturbances are activated over different periods of time.

A disturbance in F_B occurs from $t = 0$ until $t = 20$, and the composition set-point is perturbed thereafter until the end of a 40-hour simulation. Further details can be found in Appendix A.4.

Figures 8.2 and 8.3 provide high density time series plots of the standardised and Skogestad scaled simulation results respectively. Note the attenuation of outlet flow rate and height signals as well as the amplification of the composition signal in the Skogestad scaled data.

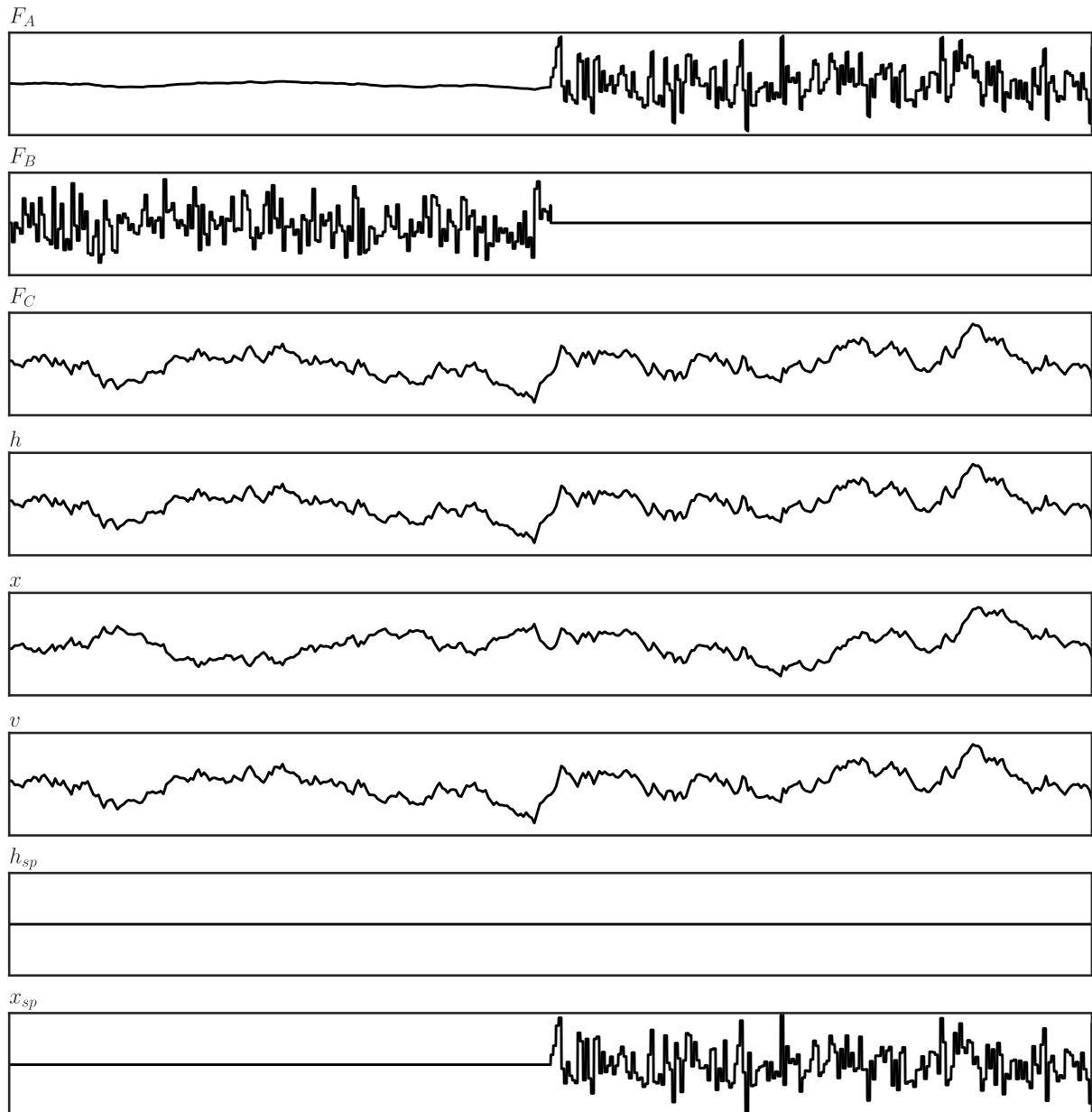


Figure 8.2: High density time series plot of disturbance [standardised]

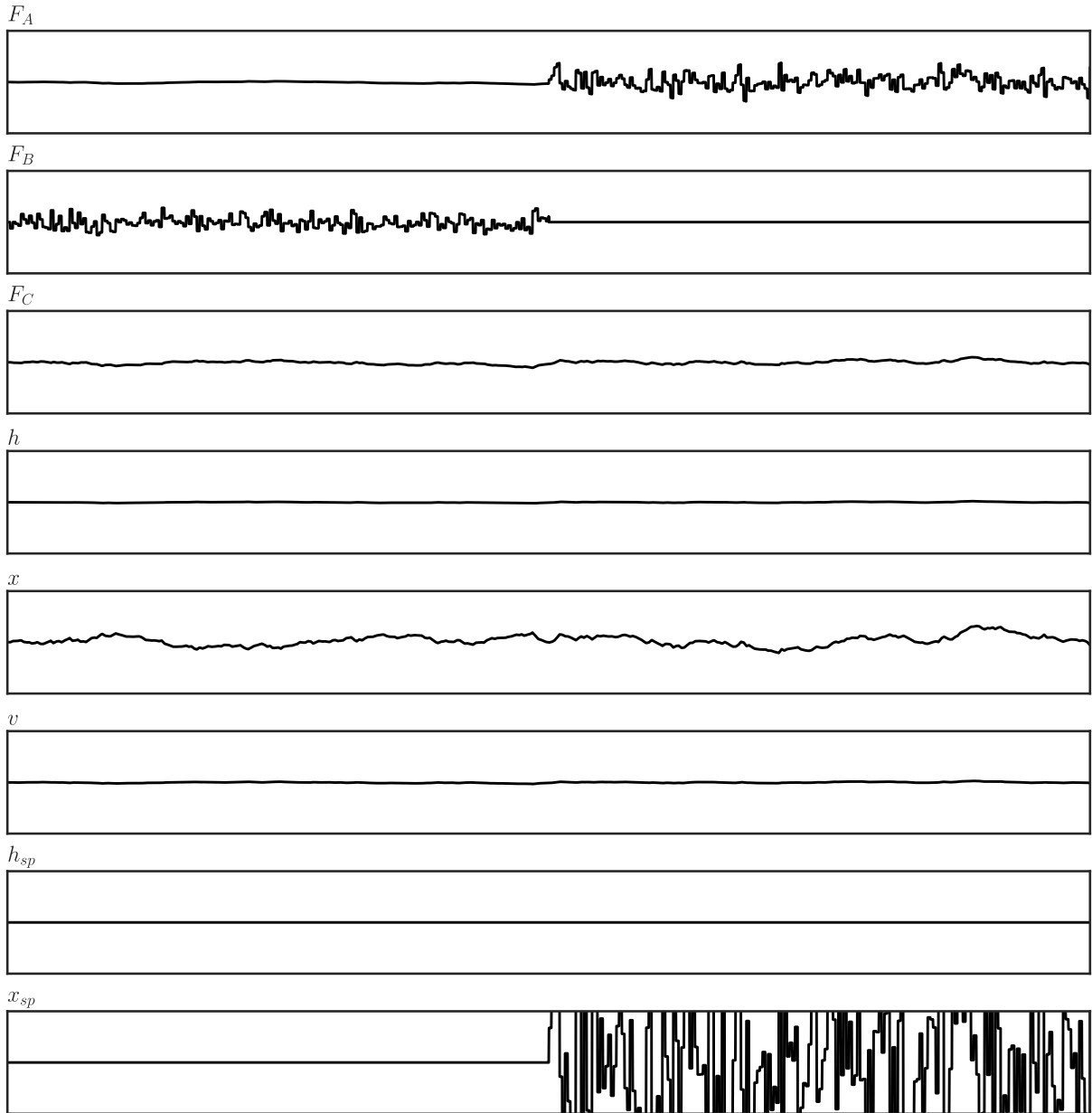


Figure 8.3: High density time series plot of disturbance [Skogestad scaled]

8.4 Network impact rankings over time

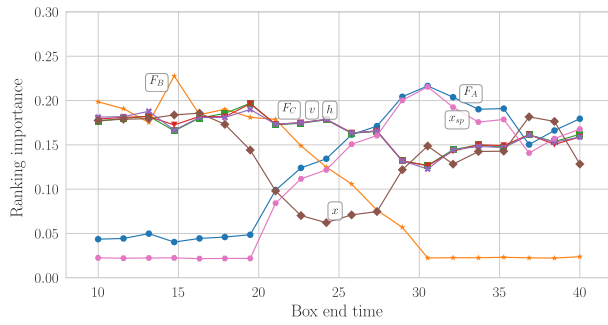
As stated in Chapter 1.2 the primary objective of the method developed in this work is to provide a data-driven, model-free process monitoring tool that can identify the most important sources of original information in a system. For the network ranking scores to become truly useful in a monitoring application, a single simple comprehensive visualisation is required, similar to common Statistical Process Control (SPC) tools such as Shewart and CUMSUM charts.

Plotting the network impact score of each variable over multiple highly overlapping time windows allows relative changes in node importance to be observed. Figures 8.4 and 8.5 show the results obtained with no significance testing of connections with and without directionality considered, respectively. Note that cross-correlation only appears in the non-directional measures plot. Figures 8.6 and 8.7 show the same results but with significance testing of connections enabled.

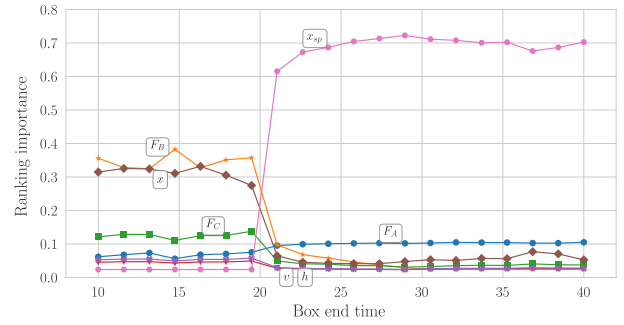
F_B should be indicated as the most important node during the first disturbance, while the expectation is that x_{sp} will dominate during the second with F_A also indicated with elevated importance. Some results follow this expectation clearly while others seem meaningless. A number of clear patterns emerge:

- Skogestad scaling consistently and significantly improved the results of both cross-correlation and kernel transfer entropy methods, validating the arguments made for its use.
- The KSG transfer entropy estimators fared better with standardisation than Skogestad scaling, but generally returned worse results than the kernel estimator with Skogestad scaling.
- The kernel estimator performed better with simple transfer entropies, while the naive KSG estimator returned better results with directional transfer entropy.
- Standardised directional KSG results are only slightly inferior to Skogestad scaled simple kernel results, indicating a viable alternative when the effort of configuring scaling limits are not considered justified.
- There is little evidence that embedding improves the KSG estimator results, and it is detrimental when directional entropies are used for reasons mentioned in Section 6.1.2.
- Significance testing is not essential for getting good results, but does not appear detrimental either. The main benefit of significance testing can therefore be argued to result in an ITN that aligns better with physical connectivity, see Section 8.5.

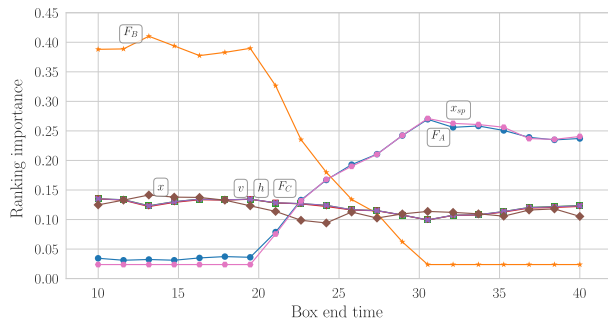
When Skogestad scaling limits are available, the simple kernel transfer entropy method should be used to infer weights on the directed graph. In the absence of proper scaling limits, the directional naive KSG estimator should be used. Significance testing is not necessary for getting good MTR node ranking plots, and may be safely skipped for fault detection, but should be enabled when the underlying ITNs are to be used for fault diagnosis.



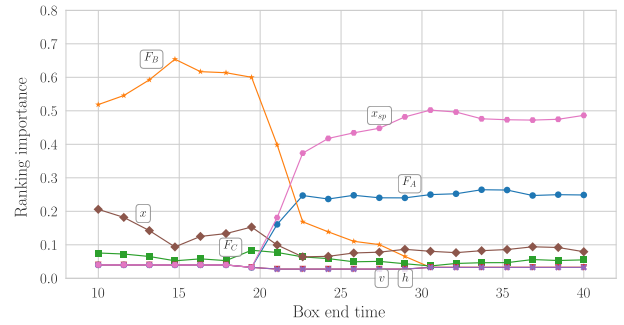
(a) cross correlation; standardised



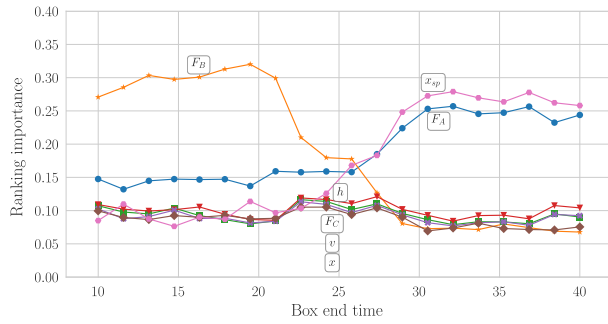
(b) cross correlation; Skogestad



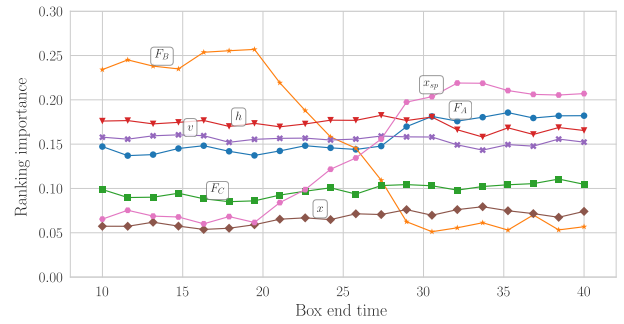
(c) TE kernel estimator; standardised



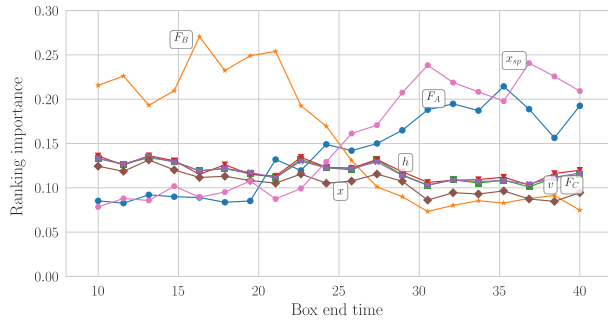
(d) TE kernel estimator; Skogestad



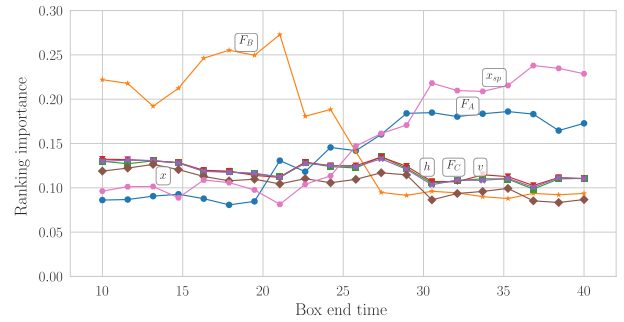
(e) TE KSG estimator; standardised



(f) TE KSG estimator; Skogestad

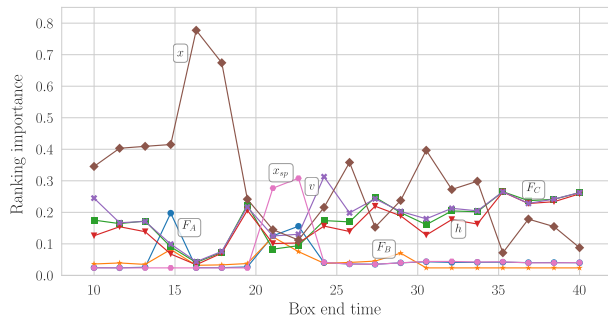


(g) TE KSG estimator; standardised; embedded

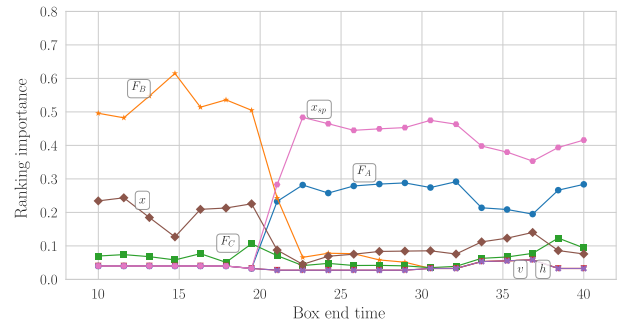


(h) TE KSG estimator; Skogestad; embedded

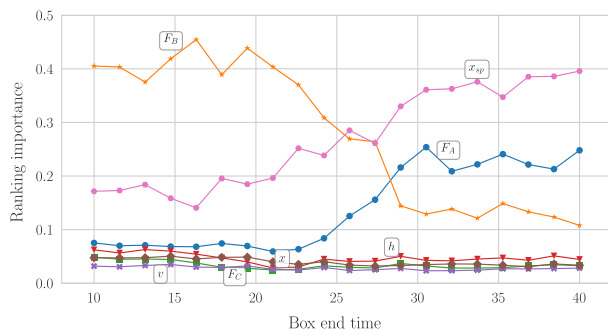
Figure 8.4: Multiple time region node rankings [simple; no significance testing]



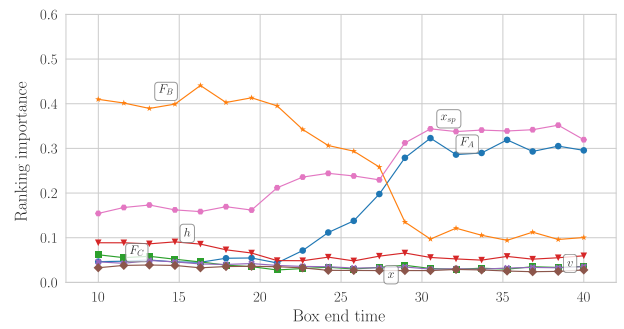
(a) TE kernel estimator; standardised



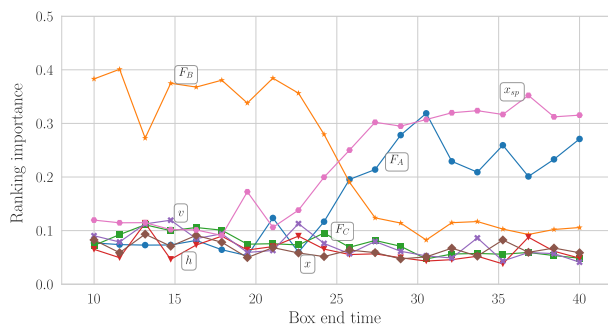
(b) TE kernel estimator; Skogestad



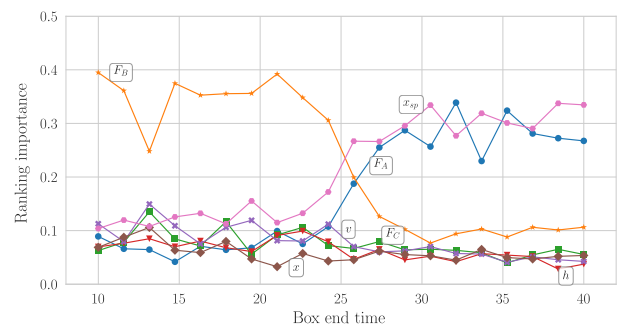
(c) TE KSG estimator; standardised



(d) TE KSG estimator; Skogestad

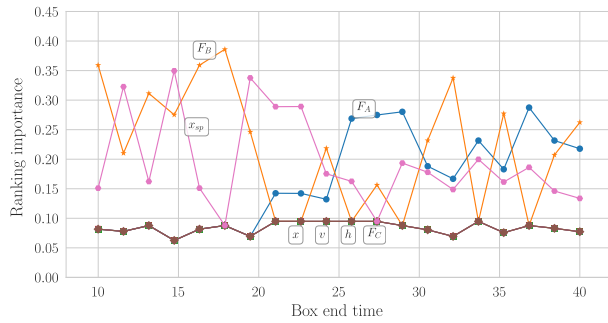


(e) TE KSG estimator; standardised; embedded

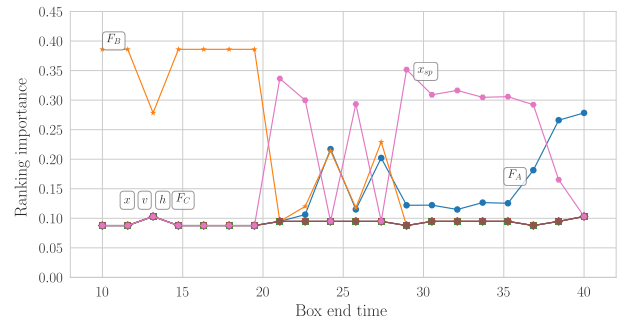


(f) TE KSG estimator; Skogestad; embedded

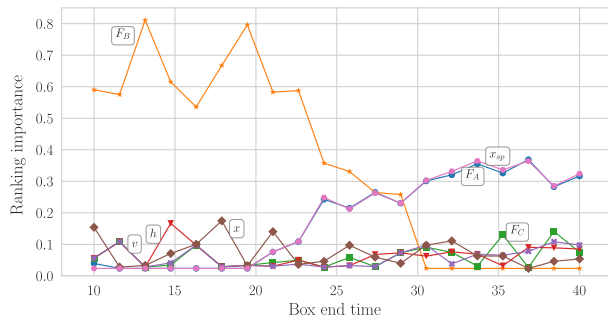
Figure 8.5: Multiple time region node rankings [directional; no significance testing]



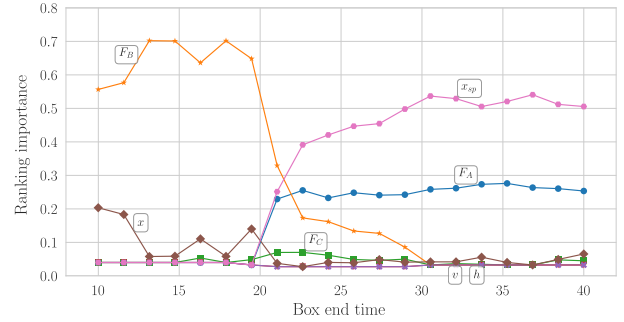
(a) cross correlation; standardised



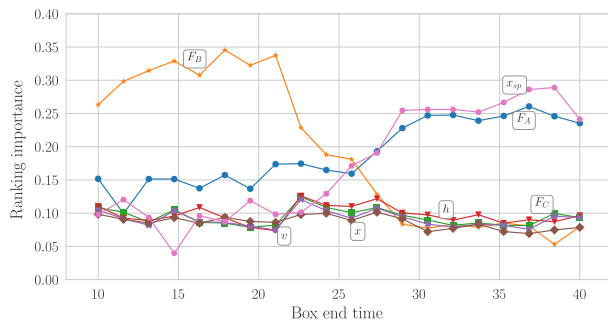
(b) cross correlation; Skogestad



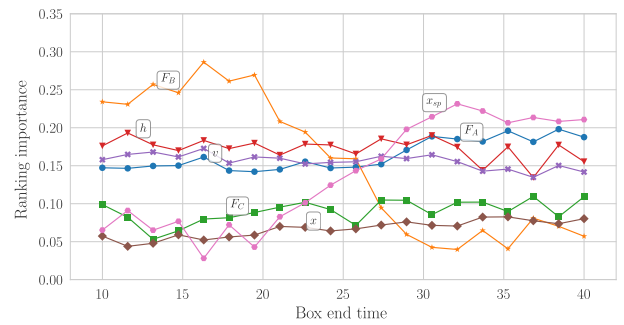
(c) TE kernel estimator; standardised



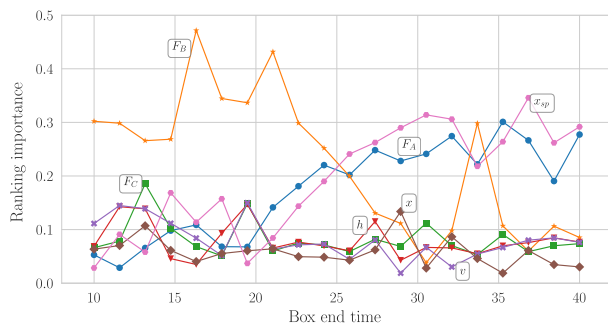
(d) TE kernel estimator; Skogestad



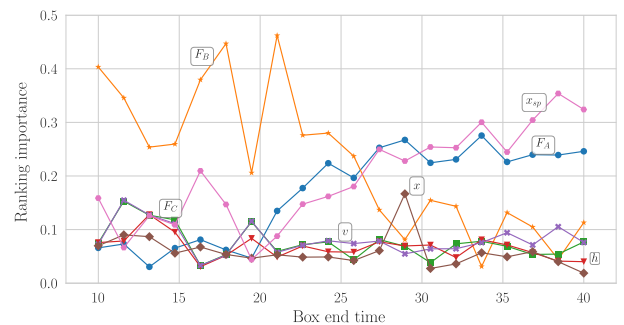
(e) TE KSG estimator; standardised



(f) TE KSG estimator; Skogestad

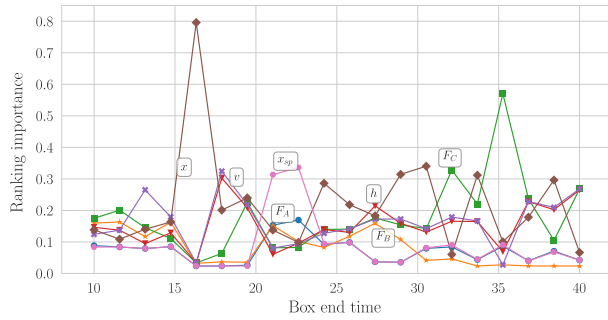


(g) TE KSG estimator; standardised; embedded

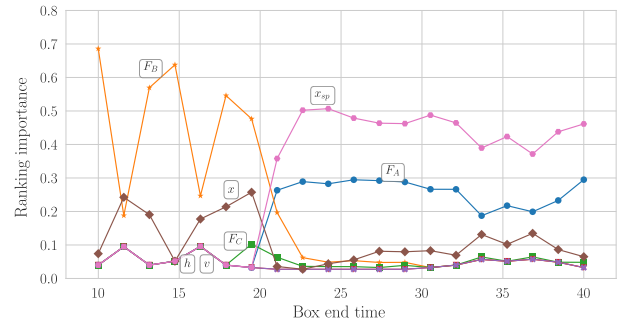


(h) TE KSG estimator; Skogestad; embedded

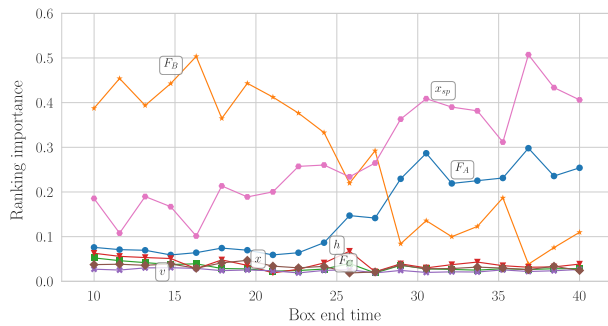
Figure 8.6: Multiple time region node rankings [simple; significance tested]



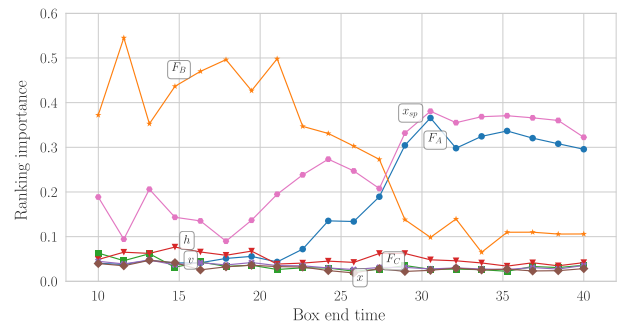
(a) TE kernel estimator; standardised



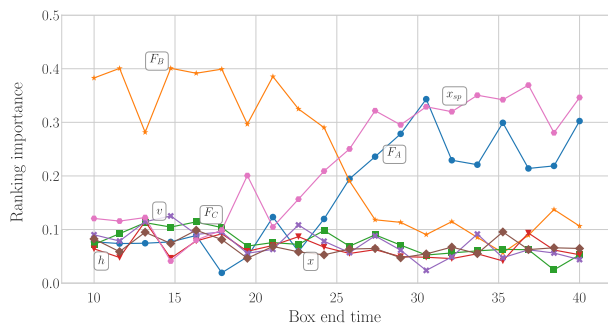
(b) TE kernel estimator; Skogestad



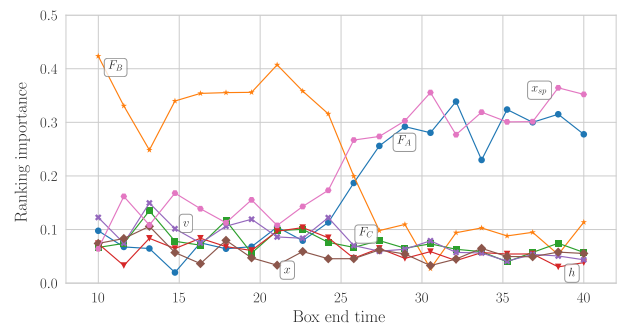
(c) TE KSG estimator; standardised



(d) TE KSG estimator; Skogestad



(e) TE KSG estimator; standardised; embedded



(f) TE KSG estimator; Skogestad; embedded

Figure 8.7: Multiple time region node rankings [directional; significance tested]

8.5 Connectivity graphs

The directed ITN for any specific time window can be used to get deeper insight into the structure of interactions. MTR network impact ranking plots are useful for fault identification, whereas ITN visualisations are more helpful in fault diagnosis.

Figures 8.8 and 8.9 shows the ITNs for the last time window investigated where the composition set-point disturbance has been active in isolation for a long time, with and without directionality considered respectively. The ITNs represent the analysis cases with significance testing applied and have been simplified according to the parameters in Table 8.2.

The results are crude at best, but contain a number of interesting features nonetheless. Note that in isolated cases involving the KSG estimator two constant signals F_B and h_{sp} have incoming or outgoing connections, indicating that they passed a significance test. This is due to the KSG estimator adding a small amount of noise to signals in order to ensure differentiation between points, a requirement for its convergence. Therefore, it is important to remove stationary signals before analysis with the KSG estimator.

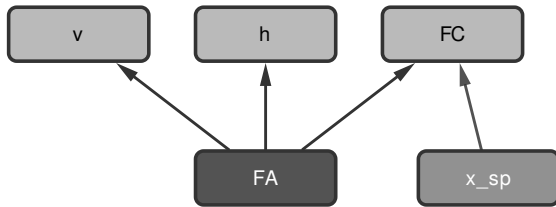
Another obvious mistake is that although x_{sp} is an independent source of information, some methods indicate incoming connections. Applying the bi-directional significance tests described in Section 4.7 might eliminate these errors. Alternatively, Section 9.5.2 discusses a procedure that can be followed to resolve spurious bi-directional couplings.

Inspection of Figure 8.6 (b) shows that although cross-correlation correctly indicated x_{sp} as the driving variable for most of the time windows, a rapid reversal to F_A occurred in the last few windows, resulting in the disappointing ITN of Figure 8.8 (b). This is likely due to a phase shift effect that negatively effected the directionality significance test.

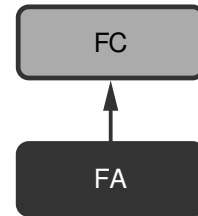
In general, the ITNs produced by the KSG estimator are a better resemblance of reality, and Skogestad scaling appears to allow for better detection of propagation paths resulting in a more hierarchical graph (compare Figure 8.9 (c) and (d), for example).

It is likely that the visualisation of certain methods were less severely affected by the rather crude approach of eliminating connections below a certain threshold, especially considering that the absolute value of connections do not carry any particular meaning in the current implementation.

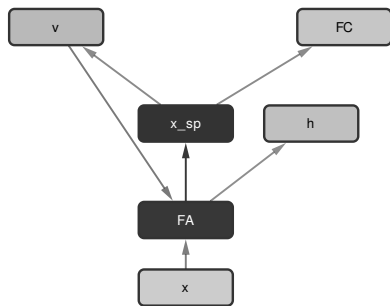
The significant contrast between the accuracy of individual ITNs and the MTR ranking results clearly demonstrate the significant benefit of performing node centrality analysis for the purpose of fault detection. Ranking does not only condense information but also filters out the effect of individual erroneous connection by aggregating over all the results. A similar result is seen in boosting models for regression and classification,



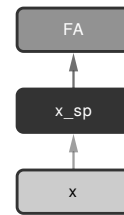
(a) cross correlation; standardised



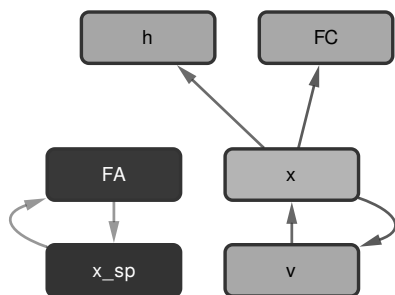
(b) cross correlation; Skogestad



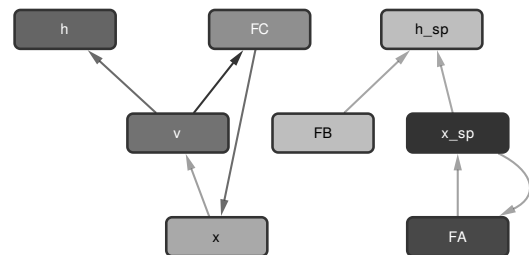
(c) TE kernel estimator; standardised



(d) TE kernel estimator; Skogestad

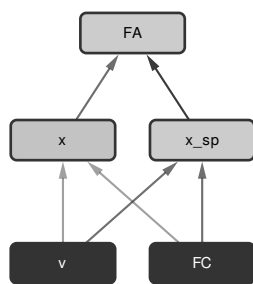


(e) TE KSG estimator; standardised

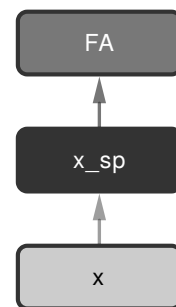


(f) TE KSG estimator; Skogestad

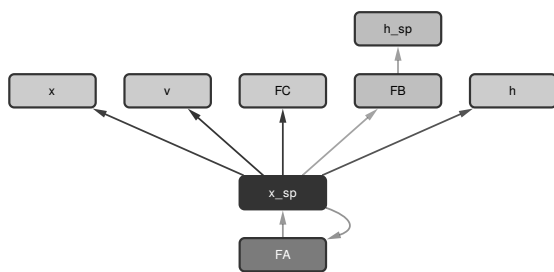
Figure 8.8: Reduced edge ITNs for x_{sp} disturbance [simple; significance tested]



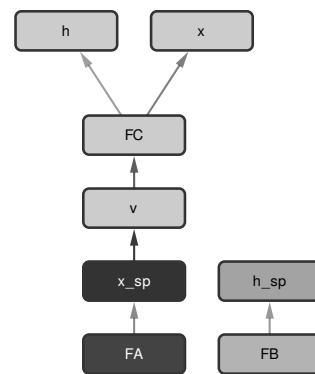
(a) TE kernel estimator; standardised



(b) TE kernel estimator; Skogestad



(c) TE KSG estimator; standardised



(d) TE KSG estimator; Skogestad

Figure 8.9: Reduced edge ITNs for x_{sp} disturbance [directional; significance tested]

where an aggregate of mediocre models manages to provide results that are better than any individual model.

Part III

Implementation

CHAPTER 9

IMPLEMENTATION AND PROCEDURE

This chapter discusses the detailed steps for performing a fault and disturbance screening analysis using the methods and software developed in this work.

9.1 Software installation

Appendix B explains the installation procedure of the accompanying software in a Linux environment. Although platform independent, that author only tested proper working in Linux.

9.2 Input files

Analysis requires time series data and configuration files that specify various parameters. An optional adjacency matrix can be defined to restrict the variable pairs (and directions) considered. Appendix C details the required file formats.

9.3 Parameter selection

To fully define an analysis case, a number of parameters need to be defined:

1. Sample size
2. Sampling rate

3. Delay optimisation search range
4. Scaling parameters
5. Time bin distribution (MTR analysis only)
6. Connectivity matrix (optional)

Some different options for obtaining the ITN are also available.

- Transfer entropy method: kernel or Kraskov
- Scaling approach: none, standardisation or Skogestad
- Significance testing model: rank order or six-sigma
- Surrogate data generation: random shuffle vs. iAAFT
- Use of bi-directional delays
- Use of an adjacency matrix
- Use of auto embedding (Kraskov method only)

Discussion of the considerations involved in selecting each of these follows in more detail below.

9.3.1 Sample size

Sample size is the number of data points involved in a single causality estimate calculation. Two primary concerns influence the selection of sample size.

Firstly, as demonstrated in Section 6.4, smaller sample sizes lead to increased variability in the results and different absolute values in the case of the KSG estimator. Complexity and stability of the process signals will dictate the minimum number of samples needed in order to estimate a representative PDF. In the test case, the signals passed through a FOPDT system. Although this model can approximate many chemical processes, highly non-linear or directional (large condition number as obtained by Singular Value Decomposition (SVD) analysis) processes will lead to a more complex relationship between the source and affected signals, resulting in a more complicated conditional PDF in need of approximation. Therefore, in certain exceptional cases, the convergence of causality measure with regards to varying sample sizes needs to be evaluated for an appropriate selection for the sample size parameter to be made.

Most process signals can be expected to have lower entropy than normally distributed noise (Gaussian distribution). It is, therefore, reasonable to argue that if the sample size is sufficient to estimate the PDF of random noise accurately, it will suffice for most process signals. However, the conditional PDFs are what is actually of importance, and may require more samples to be robust than a single signal estimation. Bauer (2005) recommended a minimum of 2 000 samples in her study of transfer entropy as a causality estimate on process time series data. In this study, measured values converged for sample sizes above 2 000, verifying this recommendation as useful in the context of this work.

The second consideration has to do with the fact that the PDF of the signals passed to the causality measure should be representative of the PDFs of signals generated by the process in general. The risk for a mismatch is the greatest when processes with long time constants are involved. Consider the case of a pure sine wave. The probability distribution of the essentially straight line segments with high gradient close to the zero crossings is significantly different from that of the areas close to the peaks and troughs. One should, therefore, take care to ensure that the time span covered by the data parsed to the causality measure calculator is of sufficient length compared to the time constants of the process involved.

Although the results presented in Section 6.1 indicated a low sensitivity of results with regards to slow time constants (at a relatively high sampling rate), a time span covering at least 10% of the dominating process time constant is recommended.

The desire to analyse an extensive period to capture an effect that occurred at an uncertain time should never motivate test size selection. Performing MTR analysis is recommended instead. This will not only be significantly faster, scaling linearly with regards to the data points covered instead of exponentially, it will also be able to more accurately pinpoint the exact time at which the disturbance or fault entered the system.

Computation time is dependent on the square of sample size as explained in Section 9.4. Since 1 000 samples will compute up to four times faster than 2 000, for example, it might be worthwhile to investigate the effect of sample size on the data that is to be analysed and compromise accordingly.

In summary, the focus should always be on using the smallest sample size that will allow for accurate estimation of the conditional PDFs involved.

9.3.2 Sampling rate

The product of sample size and sampling period define the time span covered by the analysis. Since the sample size is almost always constrained to a few thousand data points, the sampling frequency is the primary parameter that should be varied to adjust the time span incorporated into a single causality estimate calculation.

However, as shown in Section 6.5, the sampling rate affects the absolute magnitude of transfer entropy results. Therefore, since direct comparison will be complicated, analyses focusing on fast and slow dynamics should be performed in isolation.

Disturbances with faster dynamics are usually more troublesome than slower ones. However, some studies involved industrial cases where slow disturbances that result in oscillations of growing magnitude were allowed to go unchecked, eventually leading to the plant becoming unmanageable after a few days. Analyses targeting slow processes should, therefore, be scheduled periodically. A Fast Fourier Transform (FFT) analysis can be used to assist with identifying time constants of interest. The sampling rate should always satisfy the Nyquist sampling theorem for the frequency region of interest, as demonstrated in Section 6.5

It is highly unlikely that time series data captured at a resolution faster than one second is available in a plant historian. Using the suggested 2000 sample size, this translates into a time span of just over half an hour covered. However, as the time constants of massive distillation columns can be several hours, it might be useful to perform analyses on data sets sampled at lower frequencies.

9.3.3 Delay optimisation range

Analyses optimising for delay effects over a wide range at high resolution would be ideal. However, due to processing time constraints, compromises need to be made.

The range of process dead-time encountered can vary significantly. The throughput of a typical petrochemical plant is rather high compared to the hold-up volumes, and therefore the maximum time delays can be expected to be relatively small. However, this is not necessarily the case for mineral and pharmaceutical plants.

Arguments related to the trade-off between high resolution or optimisation range is similar to that of correct PDF estimation in Section 9.3.2 above under the assumption that processes with longer time delays will also have substantial dominant time constants.

As the optimising time delay is expected to be a property of the processes involved, it is not likely to change significantly under the influence of different faults and disturbances. Over time, the ideal delay range that optimises each particular directed variable pair for a specific system will become known. A typical analysis with significance testing enabled would be able to complete in a third of the time if delay optimisation over an extensive range was not needed. By detecting an appropriate time delay range for each directed variable pair and narrowing the search space accordingly, the computation load will reduce significantly. This optimisation will allow for higher-resolution calculations that will lead to more accurate results using the same resources.

The initial analyses delay optimisation search space should cover at least five minutes. If many directed variable pair weights maximise at the search space borders, the analyst should increase the delay range.

9.3.4 Scaling approach

As mentioned in Section 4.8, Skogestad scaling should always be used wherever possible. Standardising data is only a viable option if the process has been sufficiently excited in all areas and is relatively noise-free.

Skogestad scaling introduces a need for *a priori* information about the process regarding desired control bounds and input limits as well as nominal operating values. Fortunately, results are not overly sensitive to these parameters, and they are unlikely to change very rapidly.

9.4 Computation time

For fully connected systems, the calculation time is dependent on the square of process elements involved, such that a fully connected analysis on a problem with thirty elements will only take 36% of the time required to generate results for a study with fifty elements.

The calculation time is also roughly dependent on the square of samples involved in each causality estimate evaluation. Testing significance according to the rank-order method requires an additional 19 evaluation of the causality metric of choice. Calculating the significance threshold at each delay is useful for detailed analysis but slows computation significantly. The option of only performing significance testing at the time delay that maximised the causality metric is available.

The computational time required can be estimated as:

$$T = N \left(D + s \frac{C}{100 - C} \right) L^2 \frac{\tau_b}{p} \quad (9.1)$$

for the case of significance testing at the optimal delay only and

$$T = ND \left(1 + s \frac{C}{100 - C} \right) L^2 \frac{\tau_b}{p} \quad (9.2)$$

for significance testing performed at every investigated delay.

T is the total calculation time, N is the number elements in the adjacency matrix, each indicating a variable pair in a particular direction (for a fully connected adjacency matrix, N will be equal to the square of process elements), D is the number of delays to test, s

is a Boolean value indicating whether significance testing is performed, C is the required confidence in the rank-order significance test (in %), L is the number of samples, p is the number of parallel processors available, and τ_b is a benchmark time calculated by measuring the time it takes to perform a single causality estimate on 2000 samples and dividing by 2000^2 .

9.5 Recommended analysis procedures

The most efficient method of using the software tools developed in this work is likely to depend on the nature of the fault or disturbance, and how much the analyst already knows about it. Method applications can range from an original assessment of the key influences in a plant to confirmation of already existing hypotheses concerning propagation pathways. Whatever the case, in its current form, the intended use of the tool is a troubleshooting triaging aid for use by process experts and is not refined enough for drawing blind conclusions. Instead, sanity checks are required to eliminate spurious results, and visual graph interpretation is occasionally necessary to locate the actual source of information flow.

A lean approach can be used to steer future developments, with constant feedback from practising process control engineers used to refine aspects of the tool that is most relevant to practical difficulties. The current status of the work is a minimal working prototype that will facilitate the start of this process.

Different users are expected to find novel means of incorporating the tool into their existing fault diagnosis workflow. This section describes our preferred approaches for analysing a new system to detect the key variables, faults and disturbances as well as investigating the root source of a fault or disturbance starting from a specific node of interest. Process experts are expected to typically be interested in troubleshooting a specific fault with a known frequency fingerprint. Many of the principles described here apply to this scenario as well.

9.5.1 Initial screening procedure

1) Data cleaning and preparation

Not all data tags associated with a specific section of a plant will contribute useful information to the analysis. All potential data tags relating to a process under question should be filtered to reduce the size of the problem to about 50 elements. Values of constant set-points and other unchanging data vectors will not play a role in any causal relationship, and the analyst can, therefore, safely eliminate these

from the data set without any loss in fidelity. The analyst may also eliminate redundant measurements upon confirming that no sensor faults took place during the period under investigation.

Tags at the very end of the process stream not connected via any controllers or other feedback mechanisms only have to be investigated as far as they are affected, justifying suppression of the causality estimation in the outward direction. For complex processes a high-level analysis of the most important economic variables can be used to inform more detailed, focused analysis of flagged process sections.

2) Parameter selection

As mentioned in Section 9.3, some parameters need to be set based on properties of the process and data historian. The time constants of the processes or disturbances are one of the most important factors to consider, as the time window used in each causality estimate need to cover a reasonable amount of dynamics. Appropriate time windows can be guessed by visual inspection of the time series trend or by known time constants of the processes involved.

3) Calculate initial set of results

If it is desired to analyse a specific disturbance event or fault, care should be taken to ensure that the period covered by the actual calculations captures it. The application can then be used to calculate a full gain matrix as well as node ranking results and generate a weighted digraph file (in GML format).

4) Initial inspection of weighted digraph

The analyst should inspect the weighted digraph visually to get an overview of the nature of information flow among the variables of interest. Visual inspection of the weighted digraph while continually referring to a Process Flow Diagram (PFD) or basic description of the process, such as is commonly found in the operator training manual, is arguably the most useful means of interpreting the results while using it to form a hypothesis about the cause or source of faults and disturbances.

Some operations can be performed on the weighted graph to make analysis significantly easier. Suggested methods include:

1. calculating the node centrality measures and tracing the connections responsible for the highest scoring nodes back to their source,
2. only showing the most significant edges (e.g., top quartile), and
3. arranging the network according to a node hierarchy structure.

Cytoscape (Shannon et al., 2003) is a convenient, open source and user-friendly tool for viewing and interacting with large weighted networks and supports applying the analysis steps mentioned above.

The node centrality importance scores and related ranking provides the user with an ordered list of the most connected nodes and their relative importance scores. The nodes are colour graded to make it easy to process their relative locations and the areas they influence visually. This information is of limited use, especially if the indicated nodes are unexpected or no clear link between them and known faults is apparent. If only the most important edges are made visible, it becomes significantly easier to see why some nodes have high centrality scores.

Selecting an appropriate network layout is arguably the most critical factor in determining how easy it is to interpret data presented in the form of a network visually. In the context of finding the cause of a fault, it is desirable to arrange nodes from most independent to dependent. A hierarchical network layout serves this purpose. Specifically, the yFiles hierarchic network layout algorithm provides the most useful results amongst the options available in Cytoscape.

Nodes that do not have any incoming connections, but have outgoing links to a branching network of other nodes are considered to be independent sources of information flow. These are highly likely to be as close to the causes of faults or sources of disturbances as far as sensor placement and the selection of data tags included in the analysis allows to detect. On the other hand, nodes that have none or few outgoing connections, while being significantly influenced by other nodes, are not expected to be essential to consider as potential sources of faults or disturbances. However, if they are high-importance tags, such as final product quality, it is likely that the original fault was detected based on their values.

Although achieving a distinction among nodes based on connectivity is identically the purpose of the node centrality scores as calculated according to the specific scheme presented in Chapter 5, the ranking scores do not always necessarily align with the most critical causal variables. Deviations between maximum causal connectivity and the source of causal information come from many sources. For example, if a node serves as a significant branching point for a disturbance, it will connect to many other chains of nodes, some of which might be original sources of information flow due to other influences, and this might boost its importance score beyond that of the node that influenced it initially.

The hierarchical layout algorithm cannot be executed on fully connected graphs, as customarily produced if the recommended eigenvector centrality method is applied. Fewer edges result in clearer layouts, and the analyst should delete a significant percentage of the edges for visualisation purposes. For the kernel based transfer entropy measure, deleting all edges with weights smaller than 0.1 usually works well as this is just above the level traditionally assigned to streams of random data (which might have slipped past the significance testing step) or connections with

very high degree of separation in reality. In some cases, filtering out all edges with a weight below 0.2 was necessary to de-clutter the graph sufficiently to see the primary interactions clearly.

Having the nodes arranged in a hierarchical structure usually makes clusters influenced by different original information source nodes quite easy to identify, making the method applicable to multiple cases involving multiple distinct faults.

9.5.2 Detailed root source detection for hypothesis formulation

The initial inspection process described in Section 9.5.1 will typically result in a number of chained interactions that appear to be significant. Some approaches might be followed to analyse the properties of these signals to assist in forming a hypothesis about the actual source: The following steps partially address the potential pitfalls due to bi-directional couplings as well as scale insignificance introduced by data standardisation (see Section 9.3.4):

1. frequency analysis (fast Fourier transform),
2. delay assessment, and
3. magnitude assessment.

Most faults or disturbances in chemical processing plants are likely to be active in a specific frequency range. Looking at the fast Fourier transform of the time series signals plotted on a single set of axes will immediately indicate whether these signals share active frequency regions, and will also show the frequency region of the associated fault(s). If more than one frequency region is active, it is likely that this is due to the presence of two distinct disturbances. In this case, band-gap filtering the time series data and analysing each frequency region in isolation is recommended. A procedure for band-gap filtering time series data before analysis is available in the accompanying software.

If the results indicate an unexpected connection (implicating a false positive), the analyst is suggested to visually investigate the network while showing all edges to determine whether the link is possibly due to a high degree of separation interaction. If no convincing path is present, the false result is most likely due to an inappropriate scaling of data. If the variance is subtle in engineering units compared to the control limits, this spurious result was likely brought about by the fact that the data is standardised as discussed in Section 9.3.4.

This section presents a suggested process for tracing effects on a node of interest back to its cause. An automated graph reduction algorithm that follows this scheme is available in the software.

1) Compute full weighted digraph and network centrality metrics

2) Filter edges to make network size manageable

Having many small, almost insignificant edges will make the visual inspection and the elimination procedures that follow difficult and cumbersome. It is therefore desired to remove all edges that have weights below a specific threshold. However, if the threshold is set too high, meaningful connections might be eliminated. As mentioned in Section 9.5.1, a full weighted digraph calculation usually results in too many edges to allow for a hierarchical layout to be computed. One possible strategy is setting the threshold value at the minimum value that first allows computation of a hierarchical layout in a reasonable amount of time (say, less than 10 seconds). If too many edges remain, set the threshold higher. If some expected connections are missing with a threshold that leaves many other non-obvious links, it is good practice to investigate potential causes for this weak connection. Directional entropy will fail to detect a significant connection for very similar-looking time series data vectors, such as might occur between the measured and manipulated variable associated with a controller. A potential workaround is to use the simple entropy (together with the rules for resolving bi-directional couplings discussed below) to detect the actual direction of interaction among process variables, and then assign the absolute transfer entropy as the real weight on edge.

3) Select all first neighbours of node of interest and create subgraph

4) Remove all higher order connections

Removing all higher order connections is necessary to see the fault or disturbance propagation path as clearly as possible. Although the metric indicates indirect connections as weaker in some cases, this is not always true. However, the delays associated with logical indirect connections is usually larger than any of the other delays related to the direct path.

5) Resolve bi-directional couplings

Sometimes, the algorithm indicates the presence of bi-directional couplings. Two interactive controllers or an unstable controller is likely to result in these scenarios, especially if two tags are oscillating at the same frequency and half of the period of oscillation is smaller than the time delays search space.

Bi-directional couplings can usually be resolved accurately by simply retaining the connections with the highest magnitude. If there is no distinct difference, the optimising time delay matrix can be used to detect the direction in which the positive peak has the smallest optimising delay. If the cycles are rapid, looking at the optimal delays and normalised time trends may not be sufficient to locate the actual source of the disturbance in cases where the phase difference is more than

90 degrees. In these cases, a plot of directional transfer entropy over a range of delays is helpful. The direction which indicates the first significant positive peak is likely to be the correct one. Options for automatically generating such graphs for all variable pairs as well as deciding whether a causal link is valid based on directionality tests, which involve bi-directional delay testing for each variable pair, is available in the software. If one graph follows a smooth trend while the other is more scattered, the more continuous pattern is likely to resemble the correct direction of interaction.

If no obvious conclusions are evident from this graph, retain both edges; bi-directional couplings, in particular between a measured and manipulated variable associated with the same controller is entirely possible, and if the tuning happened to be overly aggressive, this coupling is also likely to be the source of the disturbance.

6) Identify root source nodes

After following this procedure, one or more nodes will typically remain with no incoming connections. These nodes indicate the measured value that is closest to the actual source of the disturbance among all first-order connections of the node of interest.

7) Expand network until procedure converges

If the path of interaction from an actual cause of the disturbance is very long, it is likely that the analysis indicates no connection between it and the original node of interest with a weight above the threshold necessary to allow for a hierarchical network layout to be computed. Therefore, there is a need to expand the search by making the previously identified source node(s) the node of interest and repeating steps 2-6. The networks obtained in this manner should be merged until convergence occurs, that is, the procedure indicates no new source nodes.

9.5.3 Hypothesis testing

Once a reasonable argument for the source of a fault or disturbance is available, the process engineer(s) should make an alteration inspired by this knowledge in an attempt to resolve the issue. A MTR trend analysis of the data as demonstrated in Section 8.4 is likely to be most useful in assessing the results of any adjustments made.

CHAPTER 10

TENNESSEE EASTMAN CASE STUDY

This chapter demonstrates the full process of implementation as well as reasoning applied to draw conclusions from the results of the proposed fault detection method by means of a case study on the Tennessee Eastman problem (Downs and Vogel, 1993). This problem is well-known within the academic process control community as a test case for various strategies dealing with control and fault detection methods.

10.1 Problem overview

A basic flow diagram of the Tennessee Eastman process is presented in Figure 10.1. Five main unit operations are involved: (1) a reactor, (2), product condenser, (3), vapour-liquid separator, (4) recycle compressor and (5) product stripper.

Four first-order, irreversible, exothermic reactions with reaction rates sensitive to temperature occur involving four reactants (A, C, D, E), two products (G, H) and a by-product (F). An inert (B) is also present in the system.

Table 10.1 provides a description of tags relevant to the discussion that follows.

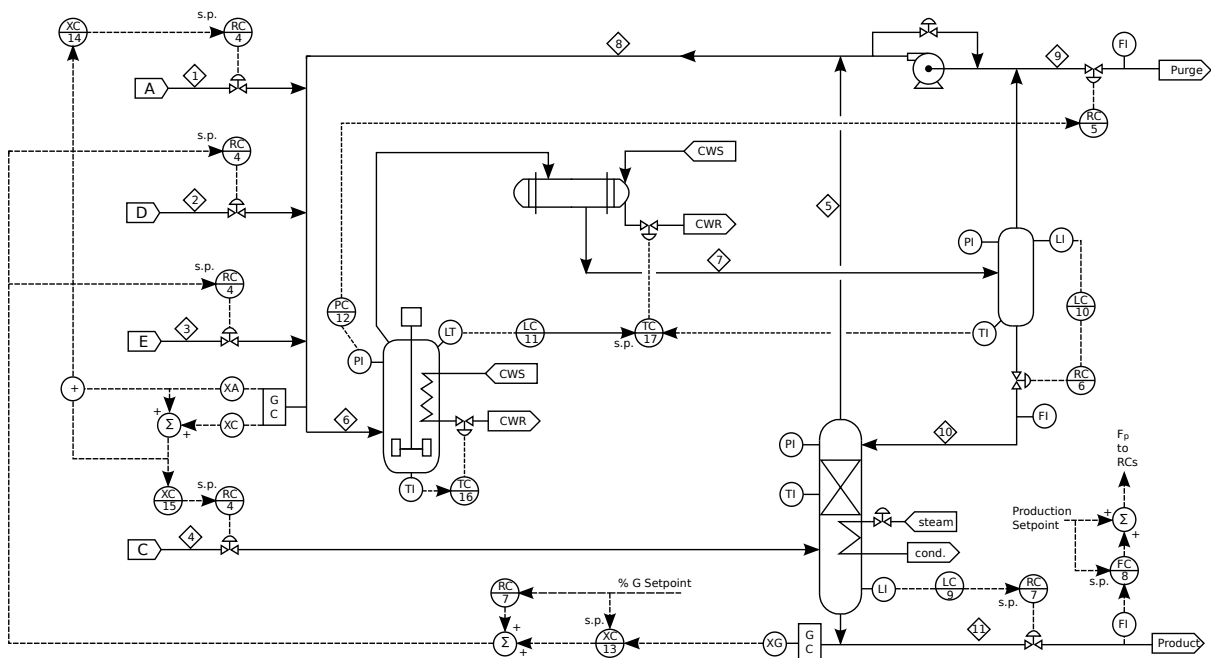


Figure 10.1: Tennessee Eastman process flow diagram

Table 10.1: Tennessee Eastman problem relevant tags

Tag	Description	Tag type
XMV 1	D feed flow (stream 2)	MV
XMV 2	E feed flow (stream 3)	MV
XMV 3	A feed flow (stream 1)	MV
XMV 4	A and C feed flow (stream 4)	MV
XMV 6	Purge valve (stream 9)	MV
XMV 7	Separator pot liquid flow (stream 10)	MV
XMV 8	Stripper liquid product flow (stream 11)	MV
XMV 10	Reactor cooling water flow	MV
XMV 11	Condenser cooling water flow	MV
XMEAS 1	A feed (stream 1)	PV
XMEAS 2	D feed (stream 2)	PV
XMEAS 3	E feed (stream 3)	PV
XMEAS 4	A and C feed (stream 4)	PV
XMEAS 5	Recycle flow (stream 8)	PV
XMEAS 6	Recycle feed rate (stream 6)	PV
XMEAS 7	Reactor pressure	CV
XMEAS 8	Reactor level	CV
XMEAS 9	Reactor temperature	CV
XMEAS 10	Purge rate (stream 9)	PV
XMEAS 11	Product separator temperature	CV
XMEAS 12	Product separator level	CV
XMEAS 13	Product separator pressure	PV
XMEAS 14	Product separator underflow (stream 10)	PV
XMEAS 15	Stripper level	CV
XMEAS 16	Stripper pressure	PV
XMEAS 17	Stripper underflow (stream 11)	CV
XMEAS 18	Stripper temperature	PV
XMEAS 19	Stripper steam flow	PV
XMEAS 20	Compressor work	PV
XMEAS 21	Reactor cooling water outlet temperature	PV
XMEAS 22	Separator cooling water outlet temperature	PV

Some base layer control strategies have been suggested in the literature. As the interest is mainly in processes under base layer control, one of these cases have been studied. In particular, the control scheme presented by Ricker (1996) was used as the basis for all tests. Downs and Vogel (1993) defined 20 predefined disturbances that could be activated on the Tennessee Eastman model. To test the MTR approach, as well as the ability of the technique to handle multiple simultaneous effects, two of the random variation disturbances were activated for specific periods of time with an overlapping window.

A random variation in the reactor cooling water (CW) inlet temperature was introduced at a simulation time of $T = 37$ hours and deactivated at $T = 145$. Random variations in the composition of feed stream E was introduced at $T = 109$ and deactivated at $T = 217$. A total of 252 hours of operation were simulated. The simulation time interval is $5E-4$ hours. Thus a total of 504 000 data points for each variable was captured.

Figures 10.2 and 10.3 presents high density time series plots of all relevant signals generated in the simulation with standardised and Skogestad scaling, respectively. As no desired control limits are known, the Skogestad scaling was performed between the 25th and 75th percentiles. Note the overlapping disturbances by comparing XMV 10 with XMV 8, for example.

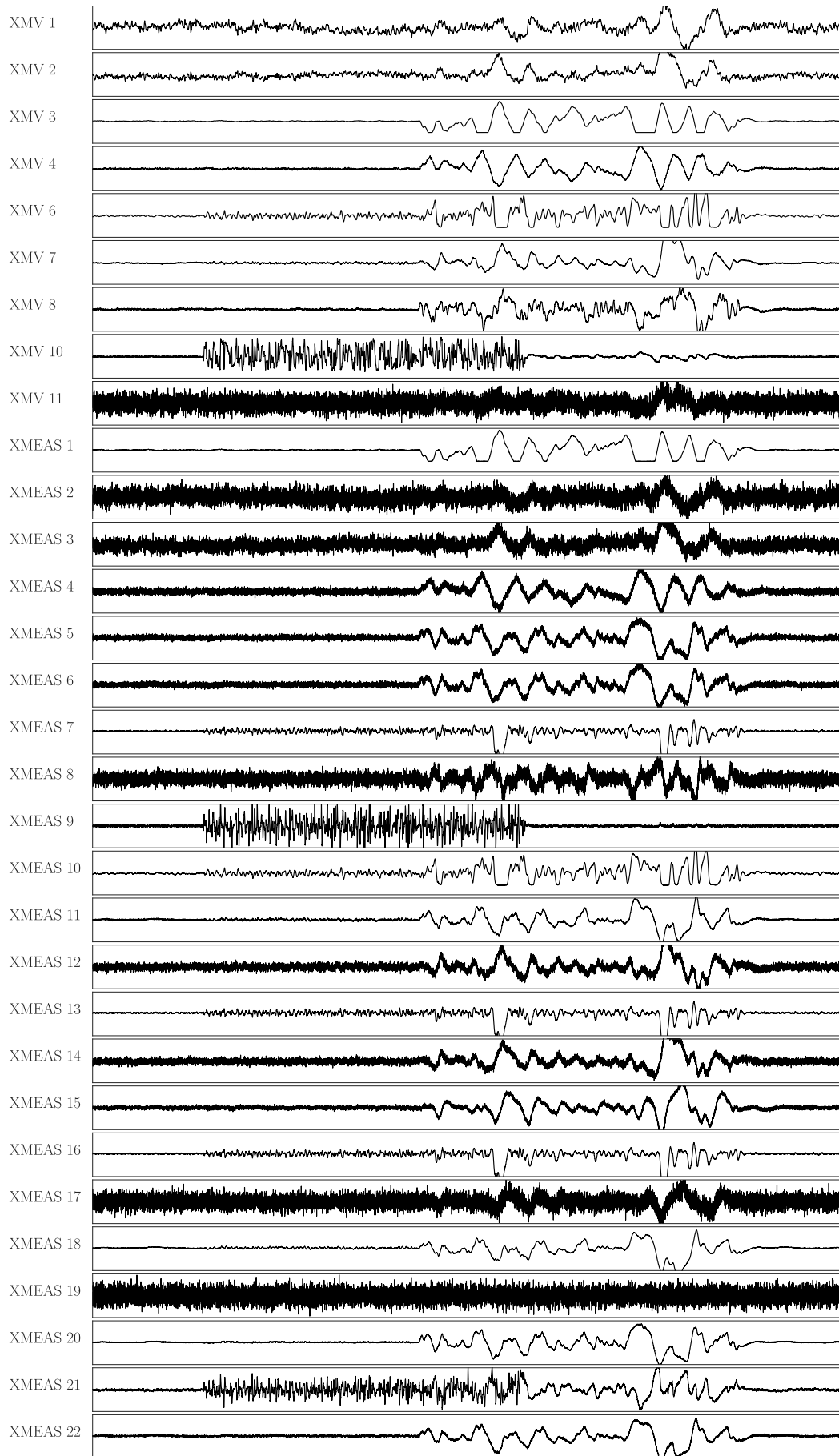


Figure 10.2: High density time series plot of disturbances [standardised]

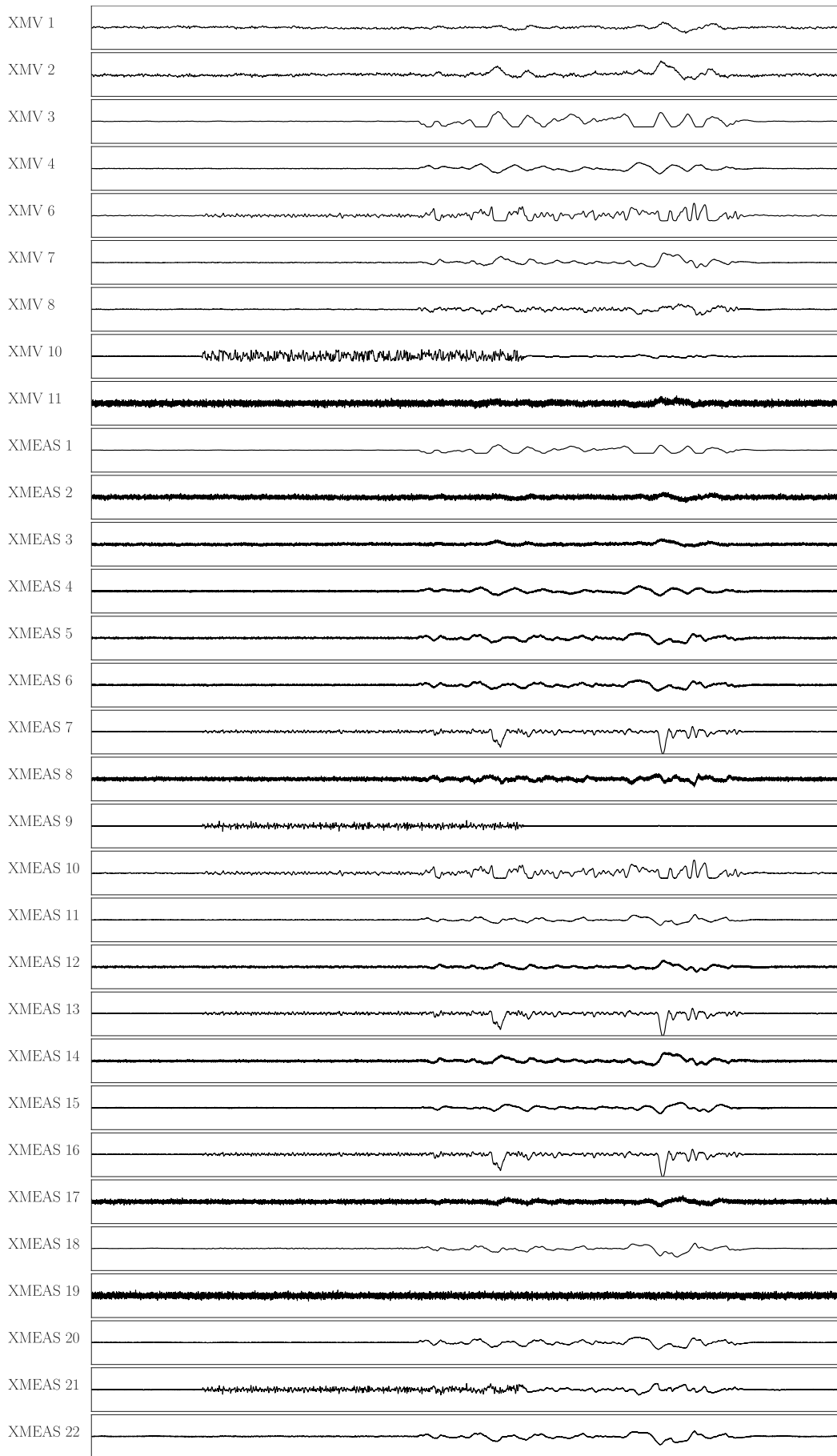


Figure 10.3: High density time series plot of disturbances [Skogestad scaled]

10.2 Choice of analysis parameters

As discussed in Section 9.3, a balance between sampling rate and sample size needs to be achieved with respect to the process time constants of interest. In the case of this problem it was decided to subsample the original data by a factor of 30, giving a final sampling interval of 0.015 hours or 54 seconds. At this resolution, if the suggested number of samples of 2000 is used, each analysis period covers 30 hours which should be enough to capture the slow dynamics of this process.

In order to achieve a 75% overlap between consecutive time windows, it was decided to spread 28 time windows covering 36 hours each evenly over the 252 hours of simulated operation. The 36 hours covered resolves to 2400 samples per time window. It was decided to optimise over 100 samples or 1.5 hours of delays in each direction, resulting in a final sample size of 2200.

The reactor cooling water inlet temperature disturbance was active in isolation during bins 2-10, while the feed composition disturbance was active in isolation for 19-27, with both disturbances present in bins 11-18.











To demonstrate the workflow and nature of results obtained, the two recommended disturbances will be investigated in isolation as well as in an overlapping time window. Finally, a transient analysis of the whole time region will be presented. It was selected to use the bins where each disturbance has been active for the longest in isolation over the full duration of the bin, that is, bins 10 and 23 for the reactor cooling water and feed composition disturbances respectively.

10.3 Reactor cooling water temperature disturbance

To investigate the effects of a reactor cooling water inlet temperature disturbance in isolation, an analysis of the results obtained for bin 10 is presented. At the start of this bin, the disturbance was already active for 35 hours and can be expected to have propagated through process units with long time constants.

The two most important unedited outputs from the developed software, namely a list of the top tags as well as directed graphs processed according to the suggestions in Section 9.5.1 is presented in Table 10.3 and Figure 10.4 below. The top ten process node importances (excluding analyser values) are also colour coded on a schematic of the Tennessee Eastman process in Figure 10.5.

Table 10.2: Transfer entropy ranking results for CW inlet temperature disturbance [kernel estimator; directional weights; Skogestad scaled; forward delays; significance tested]

Rank	Tag name	Tag description	Tag type	Relative score
1	XMEAS 21	Reactor CW outlet temperature	PV	
2	XMEAS 9	Reactor temperature	CV	
3	XMV 10	Reactor CW flow	MV	
4	XMEAS 19	Stripper steam flow	PV	
5	XMEAS 17	Stripper underflow	CV	
6	XMEAS 16	Stripper pressure	PV	
7	XMEAS 7	Reactor pressure	CV	
8	XMV 6	Purge valve	MV	
9	XMEAS 13	Product separator pressure	PV	
10	XMV 2	E feed flow	MV	

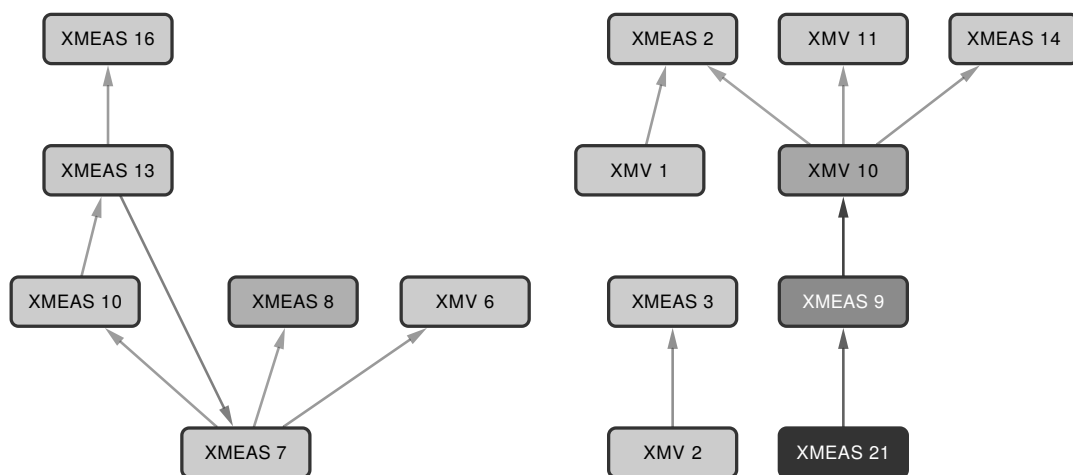


Figure 10.4: Transfer entropy information flow network for cooling water inlet temperature disturbance in Tennessee Eastman process [kernel estimator; directional weights; Skogestad scaled; bidirectional delays; significance tested]

The most significant observation that can be made from Table 10.3 is that the data tag closest to the actual source of the disturbance is given the highest importance score by a significant margin, followed by the reactor temperature and the reactor cooling water flow, which is connected to the reactor temperature via a control loop. Also note the rapid decay in importance beyond these directly related variables. Many of the other ranked variables do not feature in the reduced network structure and are likely normal operational interactions. Conditioning the ranking on weights obtained from nominal operating data will allow more of the tags involved in the disturbance causal structure

to rank higher.

By looking at the network arranged in a hierarchical layout, a very clear picture of the situation is obtained. First, note that the reactor cooling water outlet temperature is not receiving any incoming connections - indicating that it is an independent source of information flow in the system. A clear propagation path from the reactor cooling water outlet temperature to the reactor temperature, to the reactor cooling water flow MV is indicated.

The ranking results from 32 different combinations calculation choices is presented in Appendix D. Results that show more distinct relative rankings are more useful for identifying unique sources. In general Skogestad scaling performed better than standardisation, and the kernel estimator gave a more direct ranking of the most important variables. The benefit of using the difference between the forward and backward entropy as opposed to just the forward value is unclear. In the case of standardisation with the kernel estimator the difference measure gives poor results. Significance testing appears useful in most cases, and definitely helps with graph simplification for troubleshooting purposes. These observations are aligned with that of Section 8.4.

Figure 10.6 presents the results obtained with the KSG estimator. Note all the knock-on effects of the cooling water disturbance indicated, including a number of feedback loops.

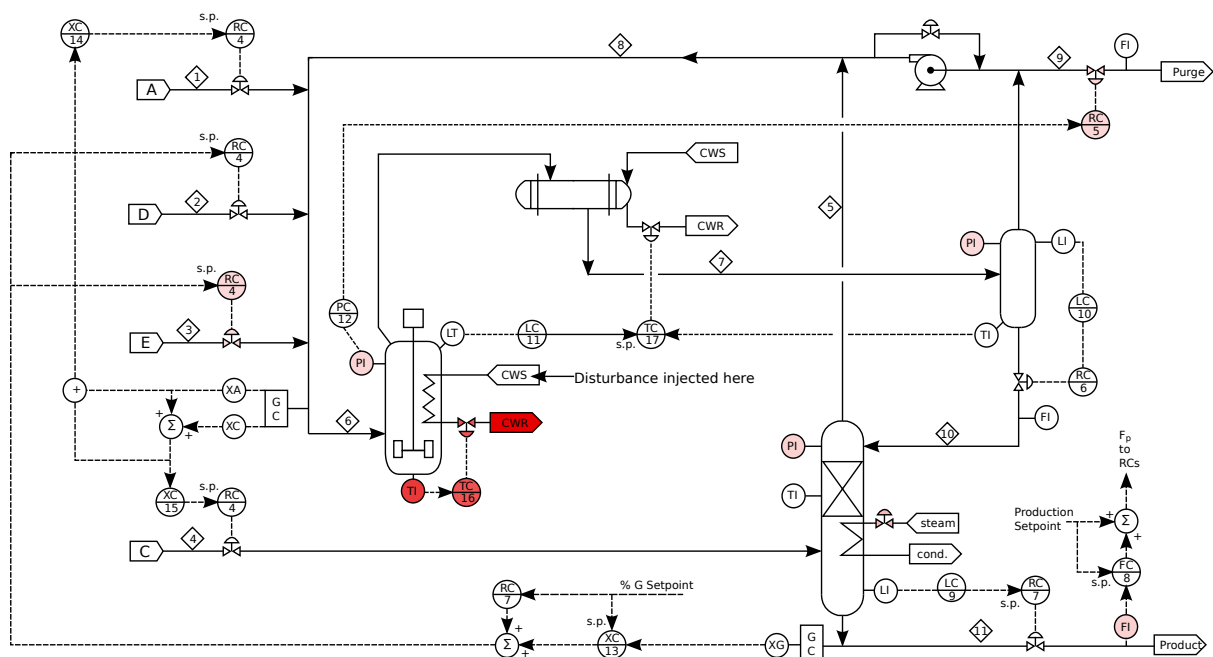












Figure 10.5: Tennessee Eastman schematic with cooling water inlet temperature disturbance effects highlighted

Cross correlation ranking results did surprisingly well. However, the causal graphs are not very useful, mostly consisting of disjointed and low-order graphs after simplification. Table 10.3 shows the ranking results using cross correlation according to Skogestad scaling with significance testing of edges performed. Note the abundance of other temperature related tags.

Table 10.3: Cross correlation ranking results for CW inlet temperature disturbance [Skogestad scaled; significance tested]











Rank	Tag name	Tag description	Tag type	Relative score
1	XMV 10	Reactor CW flow	MV	
2	XMEAS 21	Reactor CW outlet temperature	PV	
3	XMEAS 9	Reactor temperature	CV	
4	XMEAS 22	Separator CW outlet temperature	PV	
5	XMEAS 18	Stripper temperature	PV	
6	XMEAS 20	Compressor work	PV	
7	XMEAS 3	E feed flow	PV	
8	XMEAS 17	Stripper underflow	CV	
9	XMV 11	Condenser CW flow	MV	
10	XMEAS 6	Recycle feed rate	PV	

10.4 Feed composition disturbance

To investigate the effects of a feed composition disturbance in isolation, an analysis of the results obtained for bin 23 is presented.

As in the case of the reactor cooling water disturbance, a list of the top tags and a simplified directed graph is presented in Table 10.4.

Table 10.4: Transfer entropy ranking results for feed composition disturbance [kernel estimator; directional weights; Skogestad scaled; forward delays; significance tested]










Rank	Tag name	Tag description	Tag type	Relative score
1	XMEAS 22	Separator CW outlet temperature	PV	
2	XMEAS 19	Stripper steam flow	PV	
3	XMEAS 20	Compressor work	PV	
4	XMEAS 13	Product separator pressure	PV	
5	XMV 1	D feed flow	MV	
6	XMEAS 21	Reactor CW outlet temperature	PV	
7	XMV 7	Separator pot liquid flow	MV	
8	XMEAS 16	Stripper pressure	PV	
9	XMV 3	A feed flow	MV	
10	XMEAS 11	Product separator temperature	CV	

In this case, the actual source of the disturbance is not immediately apparent. It is clear that the control loops that get their setpoints from the composition controller are very active. However, note that the stripper temperature and compressor work are some of the measurements closest to the disturbance, while most of the other high-ranking nodes point to a disturbance in the composition.

Recall from the discussion in Section 9.5.1 that node centrality scores might be high for nodes that serve as distribution points for disturbances, even if they are not powerful sources of disruption themselves. Compressor work (XMEAS 20) affects all but four of the nodes in the network and demonstrates this effect. Stripper steam flow (XMEAS 19) was weakly influenced and weakly influenced almost all other nodes which results in an unfortunately high place in the ranking scores while eliminated in the graph simplification step, such that it is not included in Figure 10.7.

Table 10.4 indicates the ranking results using cross correlation according to Skogestad scaling with significance testing of edges performed. Only the first nine positions are shown as all other nodes were equally ranked according to the common reset probability score. The ranking scores do not correspond very well with those obtained using transfer entropy. From a first principles perspective the latter is considered to be more relevant, showing the benefit of a non-linear, robust causal measure.

Table 10.5: Cross correlation ranking results for feed composition disturbance [Skogestad scaled; significance tested]

Rank	Tag name	Tag description	Tag type	Relative score
1	XMEAS 9	Reactor temperature	CV	
2	XMEAS 8	Reactor level	CV	
3	XMV 6	Purge valve	MV	
4	XMEAS 6	Recycle feed rate	PV	
5	XMEAS 19	Stripper steam work	PV	
6	XMEAS 15	Stripper level	CV	
7	XMEAS 17	Stripper underflow	CV	
8	XMEAS 5	Recycle flow	PV	
9	XMV 11	Condenser CW flow	MV	

10.5 Multiple time region analysis

Figure 10.8 shows the MTR monitoring plot generated over a range spanning both the cooling water and stripper feed composition disturbances, with the most important nodes detected in bins 10 and 23 above indicated.

Note that bin 20 seems to relate to a much more significant disturbance in stripper related variables compared to bin 23 which was analysed in detail. This is in line with the time series plots in Figure 10.3, and serves to illustrate the potential importance of continuous ranking calculations in facilitating a proper understand of how a disturbance evolves through process over time, ultimately allowing for more accurate diagnosis.

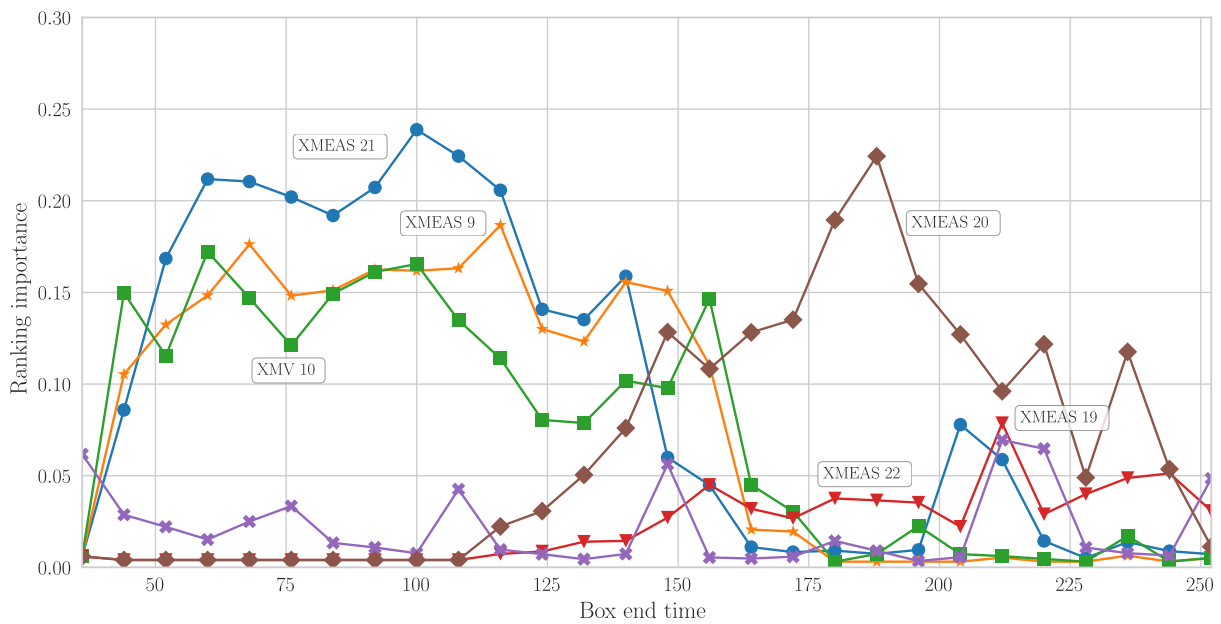


Figure 10.8: Multiple time region absolute ranking scores for periods covering both reactor CW inlet temperature and stripper feed composition disturbance in Tennessee Eastman process [kernel estimator; directional weights; Skogestad scaled; forward delays; significance tested]

CHAPTER 11

CONCLUSION

This chapter summarises the main deliverables of this work. Our impression of the degree to which the desired outcomes were achieved is mentioned and motivated.

11.1 Overview of the contributions of this work

11.1.1 Software package development

A software package referred to as FaultMap implements efficient cross-platform end-to-end execution of the proposed methods, taking only data and configuration files as input. Independent researchers have used this software package with success in manufacturing applications, including in High Performance Computing (HPC) deployments. The software was cited in a conference poster presentation (Zalmijn and Sigtermans, 2017).

In addition to the weight matrices and node rankings results, the package provides a number of additional outputs that facilitate deeper analysis, and certain classes of plots can be configured to generate automatically. Additional outputs include **GML** files for graph analysis in Cytoscape as well as causal measures and associated significance thresholds over a range of delays for every bivariate data pair in every time window investigated.

11.1.2 Motivation for applying information theory to FDD

Chapter 2 provides a brief overview of key information theory concepts and motivates the utility of information decomposition for the purpose of FDD. This chapter also discusses

the simplifications that removes the current implementation from a full multivariate information decomposition analysis, and why these are appropriate for the selected application. This work presents a thorough survey of prior art related to digraph modelling of processes (Chapter 3) and data-driven inference of process structure (Chapter 4).

11.1.3 Method and parameter selection guidelines

Chapter 6 studied the sensitivity of results to various time series data properties and developed guidelines for choosing parameters that helps to ensure successful implementations. Chapter 9 suggests a work flow for applying the method on practical industrial problems.

An extensive combination of methods and was tested, and combinations that work well for every case investigated as identified, while others returned results of very poor quality. Heuristics that indicate the most appropriate combinations is developed in Chapter 8.

11.1.4 Unique elements

As far as could be determined, the following elements have not been attempted in prior art related to the context of this work:

Scaling according to control limits

Results indicate that scaling according to control limits (referred to as Skogestad scaling in this work) provides significant benefit in certain situations.

Bi-directional delay analysis with associated significance testing logic

Directional significance testing assists in determining the true direction of interaction between coupled variables (Section 4.7.1). This is necessary for an unsupervised optimising search over a range of delays to be robust.

Use of KSG estimators and auto-embedding

The implementation of these techniques in open source code is relatively new and their application in literature mostly limited to neuro- and geoscience. The KSG estimator exhibits significant advantages in the absence of scaling according to control limits and is also better at accurately detecting long chains of interaction.

11.2 General conclusion

The presented techniques show promise in assisting a process expert to identify the source of a disturbance or fault significantly faster than by following more classical work flows,

usually limited to studying time series plots or individual loop KPIs. The methods have proven helpful in two non-trivial industrial fault diagnosis problems encountered in the petrochemical industry.

As the results come without robustness guarantees or confidence intervals and may require some effort in interpretation, they are best presented as advisories to panel operators and should not be the basis of automated action pending further development.

The concept of identifying the “true source” of a disturbance or fault in a process area that involves multiple interacting elements (especially bi-directional couplings) is ambiguous, since adjusting any number of the elements involved could be likely to improve rejection of the disturbance.

If a “fault” due to a poorly tuned controller (as opposed, for example, to a specific faulty valve), then it is likely that re-tuning most of the other controller elements involved can compensate for it. There are likely to be better and worse solutions, especially with regards to the directionality and potential ill-conditioning of the resulting system. Finding the most optimal solution to these scenarios will likely require the use of specific loop tuning tools together with expert knowledge as well as a bit of trial-and-error.

The chance of indicating the wrong element (false positive) and missing the real source (false negative) is significant when attempting to pinpoint a specific element as the singular source of a fault or disturbance. A tool tasked with screening faults and disturbances on a plant-wide scale will do better by indicating a cluster of elements involved. One potential approach to dealing with this problem is presented in Section 12.2.2. Vague, less detailed but accurate information is likely to be more useful than specific disinformation.

In our opinion the tools presented have found sufficient common ground between theory and practice to justify significant attempts at blind industrial case studies. Chapter 12 records a number of identified potential improvements to the current implementation.

CHAPTER 12

RECOMMENDED FUTURE RESEARCH

This chapter describes and motivates some potentially lucrative areas for future research identified during this work, but did not fit into the current scope and time frame.

12.1 Causality estimation and weight determination

Proper and meaningful causality estimation for assigning weights to the connections between elements is arguably the most important step in achieving useful importance scores for process elements affected by faults and disturbances. It is also the part of the process that is the most sensitive to the various parameters that need to be selected (and therefore optimised) before the calculation can commence. A study of the literature makes it clear that the application of various information-theoretic measures, some of them only defined in recent years, is still very much experimental. It is therefore likely that refinement of these methods will follow in the near future and will allow the combined approach of causality estimation and network centrality measures for fault screening to be more robust and powerful.

12.1.1 Transfer entropy history embedding dimensions

A detailed investigation into the effect of adjusting the embedding dimension of the destination variable k , as well as the source variable l , for exemplary process data, is lacking in the current literature. Section 6 mentioned difficulties associated with the instability of auto embedding results with regards to the allowed search space.

Bauer (2005) investigated the case where the $k = 0$ and $l = 2$. Setting $k = 0$ meant that only the delay shifted value of the affected variable was part involved, implying that the following calculation is performed:

$$T_{Y \rightarrow X} = \frac{1}{N} \sum_{n=1}^N \log_2 \frac{p(x_{i+h} | \mathbf{y}_i^l)}{p(x_{i+h})} \quad (12.1)$$

This formulation destroys the notion that the measure measures information transferred to the signal that was not already present in the signal itself but only measures the information shared between the affected and causal signals. One could speculate that this could make it harder to locate the source of information in a set of signals that look similar, such as interacting controllers, and thereby destroy one of the primary motivations for using this measure. On the other hand, this measure is less likely to assign very low weights between measures that appear identical to a large extent, and this may assist the ranking problem by indicating long and strongly connected paths. A detailed investigation into the trade-off between these two expected effects might be worthwhile.

Increasing the embedding dimension l of the causal variable is likely to lead to more specific results and eliminate spurious weights. Bauer (2005) suggested shifting the multiple causal time vectors by the time delay that optimises the calculation. Therefore, with $l = 2$ and time delay h , the set of causal vectors \mathbf{y}_i^l , each consisting of N samples, will be as follows (one vector per row):

$$\mathbf{y}_i^l = \begin{bmatrix} y_i & y_{i-1} & y_{i-2} & \cdots & y_{i-(N-1)} \\ y_{i-h} & y_{i-h-1} & y_{i-h-2} & \cdots & y_{i-h-(N-1)} \end{bmatrix} \quad (12.2)$$

12.1.2 Directionality and multiple variable effects

Lizier et al. (2011) reported on the expansion of the transfer entropy measure to investigate directed causal connections between multiple variables simultaneously. This measure is useful for detecting interactions where an effect is due to many elements acting together in a particular fashion, with no single element acting as a significant source of information in isolation.

A multivariate approach might be useful in the process engineering context on processes that exhibit strong directionality, where the relative movement of two or more variables is more important to affect a destination element than the movement of a single variable.

12.1.3 Power spectrum based initial screening and importance

A plant-wide fault screening will inevitably face multiple simultaneous and interacting faults, and being able to distinguish between them and their relative effects is of primary importance to prioritise maintenance tasks. Fortunately, it is likely that different faults and disturbances will influence the process elements in different frequency regions.

Screening the FFT of all signals for recurring frequency regions of high amplitude will indicate the frequency regions of interest as well as the tags associated with each. Personnel can then analyse these frequency and element sets in isolation and increased detail by band-gap filtering the data to suppress all other frequencies and performing the causal estimation method and network centrality analysis to indicate the likely path of propagation for each specific fault.

The relative highest rank of any node in each frequency region set as it appears in an analysis performed for unfiltered data on the whole plant might indicate the relative importance of faults.

12.1.4 Other measures

As already discussed in the literature overview (Section 4), publications detailing the application of old, recent and ongoing advances in statistical and information-theoretic measures to FDD are abundant. A few that are expected to provide potential refinement to the method developed in this work include:

- Directed partial correlation (Mader et al., 2010)
- Directed transfer entropy (Duan et al., 2012)
- Canonical correlation (Rahman and Choudhury, 2011)
- Principal component analysis (Thornhill et al., 2002)

12.2 Network centrality measures

Although the eigenvector and Katz centrality measures are relatively straightforward, there are additional individual concepts that can make the importance scores more accurate by considering factors other than just the causal linkages and strengths.

12.2.1 Application of additional centrality measures

As mentioned in the literature study (Section 5.3) some additional network centrality measures have been defined, and some of them have the potential for shedding more light on the shared relationships between different process elements. In particular, the betweenness centrality measure might be a useful additional indicator of the likelihood that a specific element is close to the cause if it acts as a bridge between many of the other highly ranked variables.

12.2.2 Clustering elements as disturbance and fault sources

The idea of analysing process elements according to how one region affects another region is known as clustering. Methods that differentiate between inter- and intra-cluster edges, such as the Weighted Inter-Cluster Edge Ranking (WICER) method presented by Padmanabhan et al. (2005) might be essential to consider for this application.

Interactions which appear to be significant and complicated from a global point of view are commonly found to consist of many local interactions in small groups with weak interactions among the groups. Eigenvalues of the matrix that represent networks that follow the configuration of hierarchical structures with intergroup connections show specific eigenvalue distributions. Tsubakino and Hara (2012) presented an efficient method for expressing large networks in a hierarchical structure so that its eigenvectors follow these distributions.

Most chemical plants and control engineering system problems in general belong to the category of multi-agent dynamical systems. Organising these system networks to follow a high-level hierarchical structure with intergroup connections and determining importance scores for each interconnected group individually might make analysis more accessible and precise. An adapted tool will generate fault screening reports according to smaller interconnected groups, rather than individual process elements that might occur throughout the plant.

Indicating regions of concern rather than specific elements might be more helpful to control engineers. Even though the information provided is not as precise, it has a significantly increased probability of being accurate. If a tool can reduce the search space by directing a specialists' attention to a region of, say 20 process elements out of the thousands involved in a chemical plant, it provided value.

12.2.3 Incorporating meta-data in importance score calculations

Allowing additional information about process elements such as the primary value of materials involved and previously identified faults and disturbances to be considered in the calculation of element importance scores is expected to allow for greater flexibility and relevancy in results.

The idea of biasing the eigenvector network centrality measure based on individual element parameters, for example, loop KPIs, by weighing the causal connection network with a fully connected matrix where the nodes' relative metric weights its outgoing edges was already presented in Section 5.5. However, this method might not be suitable for handling multiple information types, many of which might be boolean or integer values, such as categories or flags merely indicating a specific status, such as if a specific control loop is in manual or automatic or forms part of an APC scheme.

(Gao et al., 2011) recently developed ranking algorithms that incorporate objective function optimisation on various types of additional information sources. Incorporating these concepts may significantly enhance the capabilities of the presented method.

12.3 Background and parallel computing

The calculations involved in ranking elements of large systems require significant computational effort. To allow for an unobstructed FDD work flow in industrial practice, dedicated servers can be configured to perform calculations of causality linkage strengths between nodes for various frequency bands continuously in the background, storing results in a database for future reference.

Once the individual node causal weights are available, the various post-processing operations required to produce useful outputs such as calculating relative importance scores, detecting the most significant dynamic changes over time, clustering and tracing disturbance paths require very little computation time and can be configured to work in an interactive mode.

Calculation of the causal link estimate between any two variables in any direction is an independent process from doing the same for any other directed variable pair. As work can be divided into small packets, it is ideally suited to be processed in a distributed computing.

If the enterprise do not wish to invest in a dedicated server, it is likely that the weight database could be kept reasonably up to date by running low-priority background calculations on medium-range workstation computers without affecting the user experience of office workers involved in non-technical tasks.

12.4 Multiple user collaborative web interface

Paulonis and Cox (2003) described the implementation of a large-scale controller performance assessment system implemented at the Eastman Chemical Company. The platform shared many common goals with the work described in this document, including relatively ranking controller performance, preliminary problem diagnosis, the ability to access historical performance metrics and tracking the effects of maintenance operations. However, the assessment methods were limited to individual controller performance metrics.

Ranking control loops based on total network influence as presented in this work can significantly enhance the usefulness of such a platform as it provides the ideal framework and infrastructure for the use of interconnected performance measures by different personnel.

Upon detection of a high priority fault or disturbance personnel with the best working knowledge of that particular section might be notified and requested to form and test a hypothesis based on the various results and configurations of information made available to them. If the fault appears to span more than one operating region, a shared platform supporting data annotation will make collaboration significantly easier.

12.5 Visualisation

The value of results generated will be limited to the extent end users are willing to adopt it into their current procedures and work flows. Adoption rate is expected to depend on how intuitive and user-friendly the interface to various aspects of the results are.

There are many different potential configurations for representing results. For example, the results can be overlaid on a P&ID based layout or a hierarchical graph and might be useful for looking at the problem from different angles. It should be relatively easy to switch between presentation formats.

The user experience should allow thought patterns to proceed without interruption due to the requirement to open other applications to get to needed data or to wait for a lengthy computation to complete before the continuing with an analysis. Examples to auxiliary data that might be needed include information such as the ranking scores of the neighbouring elements, the FFT of the time signals involved, long-term nominal operating levels, and even a simple plot of time trends over the relevant period.

APPENDIX A

EXAMPLE PROCESS DESCRIPTION

This section provides the process parameters and loop tuning needed to reconstruct the level and composition control process simulation.

A.1 Process design parameters

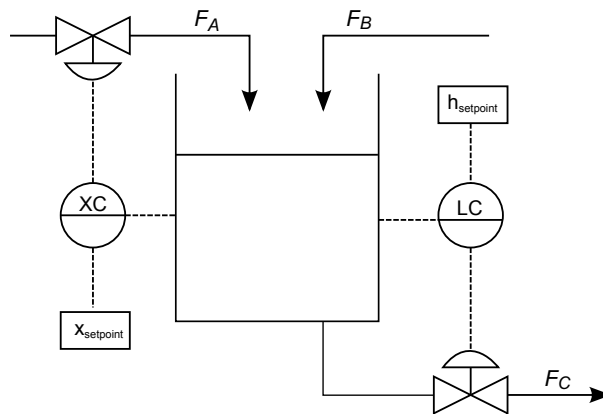


Figure A.1: Level and composition problem process flow diagram

The tank surface area is $A = 0.5 \text{ m}^2$, the density of all fluids is $\rho = 1000 \text{ kg/m}^3$.

The outlet valve follows the following relationship:

$$F_C = vk\sqrt{h} \quad (\text{A.1})$$

where v is the fractional valve opening, h is the height in the tank and $k = 200$ is the valve constant.

A.2 Control limits

The nominal operating point as well as lower and upper control limits for the process variables are presented in Table A.1. These limits are used to perform Skogestad scaling in the example results.

Table A.1: Process variable control limits

Symbol	Type	Low control limit	Nominal value	High control limit
F_A	MV	30	50	70
F_B	DV	30	50	70
F_C	PV	80	100	120
h	CV	0.8	1.0	1.2
v	MV	0.2	0.5	0.8
x	CV	0.49	0.50	0.51
h_{sp}	DV	0.9	1.0	1.1
x_{sp}	DV	0.49	0.50	0.51

A.3 Controller tuning

The Proportional Integral Derivative (PID) controllers are continuous-time and in the parallel form as shown in Equation A.2. Table A.2 shows the tuning parameters used. The anti-windup strategy used involves clamping the integrator error during saturation conditions.

$$P + I \frac{1}{s} + D \frac{N}{1 + N \frac{1}{s}} \quad (\text{A.2})$$

Error is calculated as $E = SP - PV$.

The composition loop was tuned for a robust response using a software package, while simple P-only control was implemented on the level control loop as strict control is not required there and the tank capacitance can be used to attenuate disturbances.

Table A.2: Control loop tuning parameters

Control loop	P	I	D	N	Rise time	Low limit	High limit
Composition	273	109	-48.9	0.983	5.51	0	200
Level	-2	N/A	N/A	N/A	N/A	0	1

A.4 Simulation and disturbance details

The process was simulated using an Euler fixed-step solver at an interval of 0.005 hours for a total period of 40 hours, generating 8 000 data points.

The F_B flow disturbance consists of Gaussian noise with a variance of 10 at a period of 0.1 hours added to the nominal value of 50.

The composition setpoint disturbance consists of Gaussian noise with a variance of 0.05 at a period of 0.1 hours added to the nominal value of 0.5.

APPENDIX B

SOFTWARE INSTALLATION

This section presents the full procedure for setting up the required Python distribution with associated packages in a Linux environment as well as testing the installation.

B.1 Prerequisites

The software depends on a number of libraries beyond the scope of a Python environment. These can be installed with the following commands:

```
sudo apt-get update -qq
sudo apt-get install -qq pkg-config default-jdk wget
python python-{dev,pip}
cython lib{freetype6,xft,png,xext,openblas,lapack,hdf5-serial}-dev
gfortran dvipng
sudo apt-get --no-install-recommends install texlive-base
texlive-latex-extra texlive-fonts-recommended dvipng
```

B.2 Installation and testing

The accompanying software makes use of Miniconda to create an environment with the necessary dependencies. Install Miniconda with the following commands:

```
wget http://repo.continuum.io/miniconda/Miniconda-latest-Linux-x86_64.sh
  -O miniconda.sh
bash miniconda.sh -b -p $HOME/miniconda
export PATH="$HOME/miniconda/bin:$PATH"
```

The FaultMap software can be pulled from GitHub. The distribution comes with an `environment.yml` configuration file that can be used to setup the environment. After installation you can confirm the proper functioning of the software by running the `nosetests` command:

```
git clone https://github.com/SimonStreicher/FaultMap.git
cd FaultMap
conda env create -f environment.yml
source activate faultmap
nosetests
```

APPENDIX C

CONFIGURATION, DATA AND RESULT FORMATS

This section discusses the required format for time series data, the required entries of configuration files necessary to perform analysis using the developed software, as well as the organisation of results.

C.1 Configuration files

Configuration files are used to set a number of property values which are dictated by the properties of the time series dataset involved as well as a number of general and method specific parameters that can be selected by the user. Section 9.3 discusses the process of selecting suitable values for these parameters.

Configuration files are stored as JSON dictionaries, with some of the entries being dictionaries themselves.

Different configuration files are defined for the weight calculation and node ranking steps.

C.1.1 Path configuration file

The program expects a `caseconfig.json` file in the root install directory that specifies the paths to data, configurations, outputs and the JIDT `infodynamics.jar` package used. No other manual modifications to the install directory are required.

For example:


```
{  
"dataloc": "~/faultmap/faultmap_data",  
"configloc": "~/repos/faultmapconfigs",  
"saveloc": "~/faultmap/faultmap_results",  
"infodynamicsloc": "~/repos/FaultMap/infodynamics.jar"  
}
```

C.1.2 Weight calculation configuration file

The top level of the weight calculation configuration dictionary consists of entries called `datatype`, `methods`, `scenarios` as well as various setting dictionaries and one dictionary for each scenario defined.

The `datatype` entry should be `file` if the time series data is stored in a CSV file in the data directory. The only other valid entry is `function`. This should only be used if the data is internally generated in test cases and is only included for the purpose of development and testing.

The `methods` entry is a list of strings representing the desired analysis methods. The three valid entries are:

- `cross_correlation`
- `transfer_entropy_kernel`
- `transfer_entropy_kraskov`

The `scenarios` entry is a list containing all the scenario names that need to be computed in the run. The names should be identical to the names of the scenario dictionaries defined. Each scenario needs to be defined with a distinct name.

Any number of settings dictionaries can be defined. The purpose of settings dictionaries is to define parameters that could potentially be commonly shared among various scenarios.

C.1.2.1 Settings dictionaries

Explanations of the required entries in settings dictionaries are provided below:

`use_connections`

Boolean flag entry (in JSON, this should be either a lowercase `true` or `false`) indicating whether an adjacency matrix is defined that limits the directed variable pairs that need to be considered. It is necessary for scenario dictionaries that make use of settings dictionaries with this option active to have a valid `connections` entry.

transient

Boolean flag entry indicating whether the analysis is to be performed over multiple time regions. If `true`, then the settings dictionary is required to include `boxnum` and `boxsize` entries as well.

boxnum

Integer entry specifying the number of boxes that should be evenly spaced over the available time series data. The first box will always be spaced at the beginning and the final at the end, with the remained evenly distributed over the rest of the data. If more boxes are defined than can be fitted into the available time range, they will evenly overlap.

boxsize

Integer entry indicating the desired size of each box specified in the time unit at which the data is recorded. Note that even though a box size might be set at any size that can fit into the available time data, the actual portion of the data in the box that will be analysed is dependent on the `testsize`, `startindex` and `test_delays` entries.

normalise

Boolean or string entry indicating whether the data is to be normalized before calculation, and with what method. Valid entries are `false`, `standardise` and `skogestad`. A normalisation method should always be employed for general analysis. If the `skogestad` method is chosen, the corresponding scenario dictionaries should have a valid `scalelimits` entry.

delaytype

String entry that should either be set to `datapoints` or `intervals`. If set to `datapoints` the delay interval will be equal to the sampling interval of the data after incorporating the sub-sampling interval. If set to `intervals` an additional integer entry `delay_interval` is needed that will specify how many data points should be skipped between successive delay tests. This will allow for a larger delay testing range to be covered with less tests. However, it is recommended that in most cases the `delaytype` entry be kept to `datapoints` and the delay range covered adjusted by adjusting the sub-sampling interval.

sampling_rate

A float entry indicating the sampling rate at which the data in the CSV file was recorded. This needs to be defined in the same units as the `sampling_unit` entry in order for graph axes to be scaled correctly.

sub_sampling_interval

An integer number specifying at what rate the provided time series data should be sampled for actual calculations. If the desired sampling rate is the same as that of the original rate at which the data was sampled, a value of 1 should be selected. If the number is set at N , every N^{th} entry from the data set will be picked. Note that this is not the same as mean re-sampling, which should be performed in an independent pre-processing step if desired. No effort at adjusting these values based on the properties of the neighbouring values is made.

sampling_unit

A string describing the time unit in which the original data was sampled and base of the `sampling_rate` entry. Examples include "seconds" or "hours".

testsize

An integer entry defining the number of data points used to calculate the causality measures. It is recommended that this value be kept constant at 2000 and that the time range covered be adjusted by means of the `sub_sampling_interval` entry instead.

startindex

The starting index at which the data for actual causality estimation should be sampled from inside each data box. This can be useful if it is desired to position the data selection at a specific location, or if it is known that a step has been made at a certain location and it is desired to only start sampling once the initial dynamics have died out, for example. This value is applied to the data set that is generated after the `sub_sampling_interval` has been taken into account. Note that if `bidirectional_delays` are active, `startindex` needs to be at least as high as the number of delay data points tested.

sigtest

Boolean flag indicating whether it is desired that the obtained values be subjected to significance testing. If set to true,

allthresh

Boolean flag indicating whether it is desired to have a significance threshold reported for each directed variable pair at each potential delay. This will generally take a long time to compute and is only of interest if there is some uncertainty why a specific coupling is indicated as not being significant.

C.1.2.2 Scenario dictionaries

Explanations of the required entries in scenario dictionaries are provided below:

settings

String giving the name of the settings dictionary that is to be used for analysing the scenario.

data

String containing the filename of the CSV file containing the time series data related to the scenario.

connections

String containing the filename of the appropriate adjacency matrix that is to be used for limiting tested directed pairs if the `use_connections` flag is activated in the settings.

test_delays

Integer number specifying how many delays should be tested.

causevarindexes

A list containing the indexes of variables that are considered valid cause variables for investigation. An alternative to using an adjacency matrix to limit the analysis to specific pairs. Useful if the number of connections to be analysed are small, as will typically be the case when investigating a particular area in detail when troubleshooting a previously identified fault or disturbance. Can be used in conjunction with an adjacency matrix. Note that indexes start at 0 and the order is the same as order of the columns in the time series data file. If all variables should be investigated as possible source variables, or if the selection is to be made according to a supplied adjacency matrix only, the entry should be the string `all`.

affectedvarindexes

The same as for `causevarindexes` for the affected variables.

boxindexes

List of indexes of boxes that are to be investigated for cases involving multiple time region analysis. Useful for splitting the workload among several PCs. The resulting files can be copied into the same directory and further analysis proceed as normal. If all boxes are to be calculated, the entry should be the string `all`.

bidirectional_delays

Boolean flag indicating whether delays should be tested in both directions to allow for advanced directionality significance testing.

bandgap_filtering

Boolean flag indicating whether the data should be band gap filtered before estimating causal links. Useful for investigating faults and disturbances that affect

unique frequencies in isolation. If this option is activated, the scenario dictionary need to have two additional entries: `low_freq` and `high_freq`.

low_freq

Flow that defines that low frequency cut-off for the band-gap filter, if activated. The frequency units are the inverted base time unit in terms of which the `sampling_rate` entry in the settings dictionary was defined. Band-gap filtering is performed before sub-sampling, so the filter results are not affected by altered frequency rate caused by down-sampling the signal.

high_freq

Float that defines the high frequency cut-off for the band-gap filter, if activated.

An example weight calculation configuration file for the Tennessee Eastman case is provided below. The file follows the format of `{case}_weightcalc.json` and therefore in this case is `tennessee_eastman_weightcalc.json`.

```
{ "datatype": "file",
  "methods": [
    "cross_correlation",
    "transfer_entropy_kernel"
  ],
  "scenarios": [
    "dist8_dist11_steps"
  ],
  "settings_standard": {
    "use_connections": true,
    "transient": true,
    "boxnum": 7,
    "boxsize": 5,
    "normalize": true,
    "delaytype": "datapoints",
    "sampling_rate": 5e-4,
    "sub_sampling_interval": 1,
    "sampling_unit": "hours",
    "testsize": 2000,
    "startindex": 0,
    "sigtest": true,
    "allthresh": false
  },
  "dist8_dist11_steps": {
    "connections": "connections_no_diag_ends.csv",
    "settings": "settings_standard",
    "data": "data_dist8_dist11_steps_50.csv",
    "test_delays": 50,
    "causevarindexes": "all",
    "affectedvarindexes": "all",
    "boxindexes": [5, 6],
    "bandgap_filtering": false,
    "bidirectional_delays": false
  }
}
```

C.1.2.3 Node ranking configuration file

The node ranking configuration dictionary follows a format very similar to that of the weight calculation dictionary. The node configuration dictionary allows the weight matrix files generated by the weight calculation step to be copied to the source directory for the node ranking method and used without any name changes.

The top level entries are `datatype`, `weight_methods`, `rank_methods`, `scenarios` as well as various setting dictionaries and one dictionary for each scenario defined.

The `weight_methods` refers to the name of the methods used to calculate the weight matrix in the source directory. It is used to locate the correct source file and has the same structure as the `methods` entry in the weight calculation configuration file.

The `rank_methods` entry defines the method(s) that is to be used to calculate the node rank values. It is a list of strings, with `eigenvector` and `katz` being the only two valid entries.

The `scenarios` entry is a list containing all the scenario names that need to be operated on. The names should be identical to the names of the scenario names used in the weight calculation step.

Explanations of the required entries in the settings dictionaries are provided below:

use_connections

Boolean flag entry (in JSON, this should be either a lowercase `true` or `false`) indicating whether an adjacency matrix is defined that limits the directed variable pairs that need to be considered. It is necessary for scenario dictionaries that make use of settings dictionaries with this option active to have a valid `connections` entry.

use_bias

Boolean flag entry indicating whether a bias reset probability vector should be incorporated into the node ranking procedure. Only applicable to the eigenvector method. If `true`, then the scenario dictionaries making use of the settings dictionary is required to include a valid `biasvector` entry.

dummies

Boolean flag entry indicating whether dummy nodes should be added to the network during the node ranking procedure.

Applicable entries in the scenario dictionaries include:

settings

Same as for weight calculation configuration files.

connections

Same as for weight calculation configuration files.

biasvector

String containing the filename of the appropriate bias reset probability vector that is to be used if the `use_bias` flag is active in the settings dictionary.

m The damping factor. Applies only to the eigenvector ranking method.

alpha

The attenuation factor. Applies only to the Katz ranking method.

boxindexes

List of indexes of boxes that are to be analysed. If all boxes are to be used, the entry should be the string `all`.

An example node ranking configuration file for the Tennessee Eastman case is provided below. The file follows the format of `{case}_noderank.json` and therefore in this case is `tennessee_eastman_noderank.json`.


```
{
  "datatype": "file",
  "weight_methods": [
    "transfer_entropy_kernel"
  ],
  "rank_methods": [
    "eigenvector",
    "katz"
  ],
  "scenarios": [
    "dist8_dist11_steps_biased"
  ],
  "settings_standard": {
    "use_connections": true,
    "use_bias": true,
    "dummies": false
  },
  "settings_standard_bias": {
    "use_connections": true,
    "use_bias": true,
    "dummies": false
  },
  "dist8_dist11_steps": {
    "settings": "settings_standard_bias",
    "connections": "connections_full.csv",
    "biasvector": "biasvector_pressup.csv",
    "m": 0.99,
    "alpha": 0.9
  },
  "boxindexes": [5]
}
```

C.2 Time series data

Time series data must be supplied in a comma separated (CSV) file that adheres to the following format:

Time	Tag 1	Tag 2	Tag 3	...
t_1	x_{1_1}	x_{2_1}	x_{3_1}	...
t_2	x_{1_2}	x_{2_2}	x_{3_2}	...
\vdots	\vdots	\vdots	\vdots	...
t_N	x_{1_N}	x_{2_N}	x_{3_N}	...

The time can be integer or floating point numbers. If a real valued time is desired, it is suggested that UNIX time be used as is compatible with plotting tools such as TOPCAT. Note that UNIX time does not natively support resolutions smaller than one second and entries should be restricted to the appropriate integers. Although workarounds are possible, it is not currently supported.

Note that the CSV files are converted to the H5 data format in a preprocessing step in order to allow faster execution of the data reading functions. The time series data should be stored in the data directory whose location is defined in the path configuration file under the `cases` directory.

Each case is usually associated with a different plant, or section of plant. Each case is allowed to have different “scenarios”. Scenarios may vary in any of the parameters that can be specified - the use of this structure is up to the discretion of the user. To keep organization simple, it is suggested that investigations where parameters such as the sampling interval or time delay optimization range were varied be defined as different scenarios under a common case.

C.3 Optional: Adjacency matrix

If the `use_connections` flag is activated in a settings dictionary, and accompanying filename of an adjacency matrix must be provided in the accompanying scenario configurations under the `connections` key. This adjacency matrix should adhere to the following format:

	Tag 1	Tag 2	Tag N	...
Tag 1	0	1	1	...
Tag 2	1	0	0	...
\vdots	\vdots	\vdots	\vdots	...
Tag N	1	0	1	...

where entries of 1 indicate that the connections should be included in the analysis and entries of 0 indicate that it should be ignored.

If no adjacency matrix is defined, a fully connected adjacency matrix will be used, resulting in the analysis of all possible variable pairs and directions. As discussed in Section 9.4, the computational time required is linearly dependent on the number of unique directed variable combinations tested. It might therefore be worthwhile to spend some time eliminating combinations that are not possible, such as restricting the connections of input and output elements (in the context of the set of variables investigated) to be outgoing and incoming, respectively.

C.4 Output organisation

The program will generate results in a number of directories organised by result type:

`boxdates` starting and ending times for a specific time bin

`data` the parsed input data in H5 format

`detrenddata` detrended data according to method of choice

`fftdata` FFT results for each scenario

`graphs` plotting outputs

`noderank` node ranking outputs, including ranking lists and graph files

`normdata` normalised (scaled) data according to methods of choice

`trends` weight and delay trends over multiple time bins

`weightdata` weight and delay arrays as well as aggregated results for each time bin

Each directory further organises results according to the following path:

`case/scenario/method/sigtest/embedding/boxindex`

The path elements are as follows:

`case` case name

`scenario` scenario name

`method` the weight calculation method

`cross_correlation` cross correlation

`transfer_entropy_kernel` transfer entropy according to the fixed bandwidth kernel estimator

`transfer_entropy_kraskov` transfer entropy according to the KSG estimator

`sigtest` indicates whether significance testing is performed

`sigtested` for significance testing

`nosigtest` for no significance testing

`embedding` indicates if embedding is performed in the estimator

`naive` for no embedding

`autoembedded` for embedding according to the Ragwitz criterion

`boxindex` the time bin index, e.g. `box012`

APPENDIX D

DETAILED TENNESSEE EASTMAN RANKING RESULTS

This section presents detailed ranking results for all combinations of analysis methods tested on the Tennessee Eastman reactor cooling water disturbance case study.

The following figures present results for all 36 combinations of weight calculation methods. Results with the most relevant tags to the disturbance, namely XMEAS 21, XMEAS 9 and XMV 10 in the top 3 are highlighted with light green. Results with more distinct relative rankings are more useful for identifying unique sources.

Variable	Absolute ranking	Relative ranking	Variable	Absolute ranking	Relative ranking
cc_bidirectional_standardised_nosigtest			cc_bidirectional_skogestad_nosigtest		
XMEAS 9	0.157	1.000	XMV 10	0.158	1.000
XMV 10	0.138	0.878	XMEAS 9	0.098	0.618
XMEAS 21	0.106	0.674	XMEAS 21	0.096	0.605
XMEAS 13	0.039	0.249	XMEAS 16	0.058	0.369
XMEAS 16	0.039	0.248	XMEAS 13	0.058	0.369
XMEAS 10	0.038	0.241	XMV 6	0.056	0.355
XMV 6	0.038	0.240	XMEAS 10	0.056	0.351
XMEAS 7	0.037	0.237	XMEAS 7	0.054	0.343
XMEAS 8	0.032	0.206	XMEAS 8	0.026	0.165
XMEAS 2	0.027	0.171	XMV 2	0.024	0.149
cc_bidirectional_standardised_sigtested			cc_bidirectional_skogestad_sigtested		
XMEAS 9	0.109	1.000	XMV 10	0.147	1.000
XMV 10	0.106	0.975	XMEAS 21	0.073	0.495
XMEAS 21	0.050	0.458	XMEAS 9	0.064	0.437
XMEAS 22	0.028	0.256	XMEAS 22	0.031	0.212
XMEAS 18	0.028	0.255	XMEAS 18	0.027	0.183
XMEAS 8	0.020	0.186	XMEAS 20	0.026	0.177
XMEAS 17	0.020	0.185	XMEAS 3	0.022	0.147
XMEAS 12	0.020	0.183	XMEAS 17	0.021	0.142
XMV 11	0.019	0.175	XMV 11	0.019	0.128
XMEAS 14	0.019	0.174	XMEAS 6	0.019	0.127

Figure D.1: Reactor cooling water disturbance results [cross correlation]

Variable Absolute ranking Relative ranking

tekernel_onedirectional_standardised_simple_nosigtest		
XMEAS 9	0.116	1.000
XMEAS 21	0.097	0.839
XMV 10	0.086	0.740
XMEAS 13	0.049	0.420
XMEAS 16	0.044	0.384
XMEAS 7	0.042	0.360
XMEAS 19	0.040	0.347
XMV 11	0.039	0.339
XMEAS 2	0.037	0.319
XMEAS 17	0.036	0.312

Variable Absolute ranking Relative ranking

tekernel_onedirectional_standardised_directional_nosigtest		
XMEAS 21	0.139	1.000
XMEAS 13	0.103	0.742
XMEAS 16	0.086	0.620
XMEAS 7	0.080	0.580
XMV 6	0.077	0.555
XMEAS 10	0.068	0.494
XMV 10	0.050	0.363
XMEAS 9	0.040	0.291
XMEAS 18	0.037	0.267
XMEAS 11	0.034	0.248

tekernel_onedirectional_standardised_simple_sigstest		
XMEAS 21	0.145	1.000
XMEAS 9	0.132	0.910
XMV 10	0.098	0.674
XMEAS 13	0.073	0.500
XMEAS 16	0.063	0.435
XMEAS 7	0.063	0.430
XMV 6	0.031	0.215
XMEAS 10	0.031	0.214
XMEAS 11	0.027	0.188
XMEAS 8	0.026	0.179

tekernel_onedirectional_standardised_directional_sigstest		
XMEAS 21	0.132	1.000
XMEAS 7	0.086	0.652
XMEAS 13	0.081	0.613
XMEAS 9	0.075	0.572
XMEAS 16	0.069	0.525
XMV 6	0.060	0.460
XMEAS 11	0.054	0.411
XMV 10	0.049	0.370
XMV 7	0.042	0.318
XMEAS 10	0.040	0.306

tekernel_bidirectional_standardised_simple_nosigtest		
XMEAS 9	0.116	1.000
XMEAS 21	0.097	0.839
XMV 10	0.086	0.740
XMEAS 13	0.049	0.420
XMEAS 16	0.044	0.384
XMEAS 7	0.042	0.360
XMEAS 19	0.040	0.347
XMV 11	0.039	0.339
XMEAS 2	0.037	0.319
XMEAS 17	0.036	0.312

tekernel_bidirectional_standardised_directional_nosigtest		
XMEAS 21	0.139	1.000
XMEAS 13	0.103	0.742
XMEAS 16	0.086	0.620
XMEAS 7	0.080	0.580
XMV 6	0.077	0.555
XMEAS 10	0.068	0.494
XMV 10	0.050	0.363
XMEAS 9	0.040	0.291
XMEAS 18	0.037	0.267
XMEAS 11	0.034	0.248

tekernel_bidirectional_standardised_simple_sigstest		
XMEAS 21	0.125	1.000
XMEAS 9	0.115	0.920
XMV 10	0.110	0.878
XMEAS 13	0.080	0.639
XMEAS 16	0.073	0.582
XMEAS 7	0.067	0.535
XMEAS 10	0.036	0.284
XMV 6	0.034	0.274
XMEAS 11	0.028	0.226
XMEAS 18	0.021	0.171

tekernel_bidirectional_standardised_directional_sigstest		
XMEAS 21	0.139	1.000
XMEAS 13	0.096	0.689
XMEAS 16	0.086	0.621
XMEAS 9	0.080	0.575
XMEAS 7	0.061	0.439
XMEAS 10	0.059	0.425
XMEAS 11	0.053	0.380
XMV 10	0.047	0.338
XMV 6	0.047	0.337
XMV 7	0.042	0.303

Figure D.2: Reactor cooling water disturbance results [kernel estimator; standardised]

Variable	Absolute ranking	Relative ranking	Variable	Absolute ranking	Relative ranking
tekernel_onedirectional_skogestad_simple_nosigtest			tekernel_onedirectional_skogestad_directional_nosigtest		
XMEAS 21	0.202	1.000	XMEAS 21	0.203	1.000
XMV 10	0.167	0.825	XMEAS 9	0.149	0.735
XMEAS 9	0.163	0.807	XMV 10	0.125	0.615
XMEAS 13	0.039	0.192	XMEAS 13	0.038	0.188
XMV 6	0.037	0.182	XMV 6	0.037	0.184
XMEAS 16	0.035	0.174	XMEAS 16	0.035	0.174
XMEAS 10	0.033	0.163	XMEAS 19	0.034	0.167
XMEAS 7	0.029	0.141	XMEAS 10	0.033	0.165
XMV 2	0.025	0.124	XMEAS 8	0.031	0.153
XMEAS 2	0.021	0.102	XMEAS 17	0.028	0.136
tekernel_onedirectional_skogestad_simple_sigtested			tekernel_onedirectional_skogestad_directional_sigtested		
XMEAS 21	0.229	1.000	XMEAS 21	0.194	1.000
XMV 10	0.186	0.810	XMEAS 9	0.150	0.775
XMEAS 9	0.181	0.789	XMV 10	0.132	0.679
XMV 6	0.030	0.131	XMEAS 19	0.054	0.278
XMEAS 16	0.030	0.129	XMEAS 17	0.036	0.187
XMV 2	0.026	0.115	XMEAS 16	0.034	0.175
XMEAS 10	0.026	0.113	XMEAS 7	0.032	0.165
XMEAS 7	0.026	0.112	XMV 6	0.031	0.159
XMEAS 13	0.023	0.100	XMEAS 13	0.030	0.155
XMEAS 8	0.020	0.085	XMV 2	0.030	0.153
tekernel_bidirectional_skogestad_simple_nosigtest			tekernel_bidirectional_skogestad_directional_nosigtest		
XMEAS 21	0.202	1.000	XMEAS 21	0.203	1.000
XMV 10	0.167	0.825	XMEAS 9	0.149	0.735
XMEAS 9	0.163	0.807	XMV 10	0.125	0.615
XMEAS 13	0.039	0.192	XMEAS 13	0.038	0.188
XMV 6	0.037	0.182	XMV 6	0.037	0.184
XMEAS 16	0.035	0.174	XMEAS 16	0.035	0.174
XMEAS 10	0.033	0.163	XMEAS 19	0.034	0.167
XMEAS 7	0.029	0.141	XMEAS 10	0.033	0.165
XMV 2	0.025	0.124	XMEAS 8	0.031	0.153
XMEAS 2	0.021	0.102	XMEAS 17	0.028	0.136
tekernel_bidirectional_skogestad_simple_sigtested			tekernel_bidirectional_skogestad_directional_sigtested		
XMEAS 21	0.232	1.000	XMEAS 21	0.189	1.000
XMV 10	0.197	0.848	XMEAS 9	0.102	0.539
XMEAS 9	0.185	0.799	XMEAS 19	0.092	0.487
XMV 2	0.028	0.121	XMV 10	0.074	0.393
XMV 6	0.025	0.107	XMEAS 8	0.066	0.352
XMEAS 13	0.024	0.102	XMEAS 13	0.041	0.216
XMEAS 2	0.023	0.100	XMEAS 17	0.037	0.195
XMEAS 16	0.021	0.089	XMEAS 3	0.036	0.189
XMEAS 8	0.019	0.082	XMEAS 14	0.031	0.164
XMEAS 10	0.016	0.067	XMV 2	0.030	0.161

Figure D.3: Reactor cooling water disturbance ranking results [kernel estimator; Skogestad scaled]

Variable	Absolute ranking	Relative ranking	Variable	Absolute ranking	Relative ranking
teksg_onedirectional_standardised_simple_nosigtest			teksg_onedirectional_standardised_directional_nosigtest		
XMEAS 21	0.053	1.000	XMEAS 22	0.055	1.000
XMEAS 13	0.052	0.996	XMEAS 11	0.050	0.899
XMEAS 16	0.048	0.909	XMEAS 13	0.044	0.804
XMEAS 7	0.047	0.889	XMV 7	0.043	0.784
XMV 10	0.044	0.840	XMV 3	0.040	0.730
XMEAS 9	0.044	0.838	XMEAS 20	0.040	0.722
XMEAS 22	0.044	0.826	XMEAS 18	0.038	0.692
XMEAS 11	0.043	0.821	XMEAS 21	0.037	0.676
XMV 7	0.039	0.745	XMEAS 16	0.036	0.648
XMEAS 20	0.038	0.727	XMEAS 7	0.035	0.630
teksg_onedirectional_standardised_simple_sigstested			teksg_onedirectional_standardised_directional_sigstested		
XMEAS 21	0.066	1.000	XMEAS 22	0.068	1.000
XMEAS 9	0.058	0.878	XMEAS 21	0.049	0.721
XMEAS 13	0.054	0.824	XMEAS 11	0.047	0.693
XMEAS 16	0.050	0.759	XMV 7	0.044	0.647
XMV 10	0.049	0.754	XMEAS 20	0.043	0.636
XMEAS 22	0.047	0.713	XMEAS 9	0.040	0.580
XMEAS 11	0.045	0.690	XMEAS 13	0.038	0.562
XMEAS 7	0.042	0.634	XMEAS 12	0.034	0.502
XMEAS 20	0.042	0.633	XMEAS 14	0.033	0.480
XMV 7	0.036	0.547	XMEAS 5	0.032	0.466
teksg_bidirectional_standardised_simple_nosigtest			teksg_bidirectional_standardised_directional_nosigtest		
XMEAS 21	0.053	1.000	XMEAS 22	0.055	1.000
XMEAS 13	0.053	0.998	XMEAS 11	0.049	0.891
XMEAS 16	0.048	0.911	XMEAS 13	0.045	0.823
XMEAS 7	0.047	0.887	XMV 7	0.043	0.775
XMV 10	0.044	0.832	XMEAS 20	0.040	0.731
XMEAS 9	0.044	0.829	XMV 3	0.040	0.718
XMEAS 22	0.044	0.821	XMEAS 21	0.038	0.689
XMEAS 11	0.043	0.807	XMEAS 18	0.038	0.682
XMV 7	0.039	0.729	XMEAS 16	0.035	0.641
XMEAS 20	0.039	0.725	XMEAS 7	0.034	0.619
teksg_bidirectional_standardised_simple_sigstested			teksg_bidirectional_standardised_directional_sigstested		
XMV 10	0.060	1.000	XMEAS 22	0.066	1.000
XMEAS 9	0.057	0.943	XMEAS 21	0.050	0.748
XMEAS 21	0.053	0.877	XMV 7	0.044	0.671
XMEAS 22	0.051	0.843	XMEAS 11	0.044	0.659
XMEAS 13	0.050	0.826	XMEAS 9	0.040	0.611
XMEAS 20	0.046	0.769	XMEAS 14	0.039	0.593
XMEAS 16	0.043	0.717	XMEAS 13	0.039	0.584
XMEAS 7	0.043	0.716	XMEAS 20	0.036	0.536
XMEAS 11	0.040	0.661	XMV 11	0.033	0.504
XMV 7	0.037	0.614	XMV 10	0.033	0.501

Figure D.4: Reactor cooling water disturbance ranking results [KSG estimator; standardised]

Variable	Absolute ranking	Relative ranking	Variable	Absolute ranking	Relative ranking
teksg_onedirectional_skogestad_simple_nosigtest			teksg_onedirectional_skogestad_directional_nosigtest		
XMEAS 21	0.054	1.000	XMEAS 22	0.052	1.000
XMEAS 13	0.050	0.938	XMEAS 11	0.048	0.922
XMEAS 9	0.049	0.921	XMV 7	0.043	0.817
XMEAS 7	0.045	0.841	XMEAS 21	0.041	0.790
XMEAS 16	0.045	0.839	XMEAS 13	0.041	0.790
XMEAS 22	0.043	0.807	XMEAS 20	0.039	0.751
XMEAS 11	0.043	0.804	XMEAS 18	0.037	0.709
XMV 10	0.043	0.796	XMV 3	0.035	0.664
XMV 7	0.039	0.722	XMEAS 9	0.033	0.635
XMEAS 20	0.038	0.707	XMEAS 7	0.033	0.629
teksg_onedirectional_skogestad_simple_sigstested			teksg_onedirectional_skogestad_directional_sigstested		
XMEAS 21	0.068	1.000	XMEAS 22	0.065	1.000
XMEAS 9	0.064	0.949	XMEAS 11	0.056	0.860
XMEAS 22	0.053	0.785	XMEAS 21	0.056	0.857
XMEAS 13	0.053	0.785	XMV 7	0.048	0.741
XMV 10	0.051	0.748	XMEAS 20	0.047	0.731
XMEAS 16	0.050	0.740	XMEAS 9	0.040	0.613
XMEAS 7	0.048	0.707	XMV 11	0.035	0.540
XMEAS 11	0.046	0.686	XMEAS 13	0.033	0.515
XMEAS 20	0.043	0.638	XMEAS 12	0.033	0.502
XMV 7	0.039	0.576	XMV 4	0.031	0.477
teksg_bidirectional_skogestad_simple_nosigtest			teksg_bidirectional_skogestad_directional_nosigtest		
XMEAS 21	0.054	1.000	XMEAS 22	0.053	1.000
XMEAS 13	0.050	0.931	XMEAS 11	0.048	0.916
XMEAS 9	0.049	0.909	XMV 7	0.042	0.800
XMEAS 7	0.046	0.851	XMEAS 13	0.042	0.789
XMEAS 16	0.045	0.844	XMEAS 21	0.041	0.781
XMEAS 22	0.044	0.821	XMEAS 20	0.039	0.747
XMEAS 11	0.043	0.803	XMEAS 18	0.038	0.713
XMV 10	0.043	0.799	XMV 3	0.034	0.653
XMV 7	0.039	0.724	XMV 4	0.034	0.636
XMEAS 20	0.038	0.709	XMEAS 7	0.033	0.627
teksg_bidirectional_skogestad_simple_sigstested			teksg_bidirectional_skogestad_directional_sigstested		
XMEAS 21	0.060	1.000	XMEAS 22	0.056	1.000
XMEAS 22	0.052	0.877	XMEAS 21	0.051	0.918
XMV 10	0.052	0.877	XMV 7	0.049	0.882
XMEAS 9	0.051	0.860	XMEAS 11	0.048	0.858
XMEAS 13	0.048	0.804	XMEAS 12	0.048	0.853
XMEAS 16	0.047	0.791	XMEAS 20	0.046	0.817
XMV 7	0.045	0.761	XMEAS 9	0.035	0.630
XMEAS 20	0.044	0.735	XMV 4	0.034	0.614
XMEAS 11	0.041	0.681	XMV 8	0.032	0.581
XMEAS 7	0.037	0.613	XMEAS 5	0.032	0.578

Figure D.5: Reactor cooling water disturbance ranking results [KSG estimator; Skogestad scaled]

BIBLIOGRAPHY

- Bauer, M. (2005), Data-driven methods for process analysis, PhD thesis, University College London.
- Bauer, M., Cox, J. W., Caveness, M. H., Downs, J. J., Thornhill, N. F. and Member, S. (2007), ‘Finding the Direction of Disturbance Propagation in a Chemical Process Using Transfer Entropy’, *IEEE Transactions on Control Systems Technology* **15**(1), 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**(7-8), 707–719.
- Bauer, M., Thornhill, N. and Meaburn, A. (2005), Cause and effect analysis of chemical processes analysis of a plant-wide disturbance, in ‘IEE Seminar on Control Loop Assessment and Diagnosis’, Vol. 4.
- Bryan, K. and Leise, T. (2006), ‘The \$25,000,000,000 eigenvector: The linear algebra behind Google’, *Siam Review* **48**(3), 569–581.
- Castaño Arranz, M. and Birk, W. (2012), ‘New methods for interaction analysis of complex processes using weighted graphs’, *Journal of Process Control* **22**(1), 280–295.
- Chen, J. and Howell, J. (2001), ‘A self-validating control system based approach to plant fault detection and diagnosis’, *Computers & Chemical Engineering* **25**(2-3), 337–358.
- 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**(2), 159–178.
- Cover, T. M. and Thomas, J. A. (2006), *Elements of Information Theory*, 2nd edn, John Wiley & Sons, Inc.

- DeDeo, S., Hawkins, R., Klingenstein, S. and Hitchcock, T. (2013), ‘Bootstrap Methods for the Empirical Study of Decision-Making and Information Flows in Social Systems’, *Entropy* **15**(6), 2246–2276.
- Díaz, I. and Rodríguez, M. (2014), On-line Fault Diagnosis by Combining Functional and Dynamic Modelling of Chemical Plants, in ‘24th European Symposium on Computer Aided Process Engineering’, pp. 679–684.
- Downs, J. and Vogel, E. (1993), ‘A plant-wide industrial process control problem’, *Computers & Chemical Engineering* **17**(3), 245–255.
- Duan, P., Yang, F., Chen, T. and Shah, S. L. (2012), Detection of direct causality based on process data, in ‘2012 American Control Conference’, pp. 3522–3527.
- Duan, P., Yang, F., Chen, T. and Shah, S. L. (2013), ‘Direct causality detection via the transfer entropy approach’, *IEEE Transactions on Control Systems Technology* **21**(6), 2052–2066.
- Duan, P., Yang, F., Shah, S. L. and Chen, T. (2015), ‘Transfer Zero-Entropy and Its Application for Capturing Cause and Effect Relationship Between Variables’, *IEEE Transactions on Control Systems Technology* **23**(3), 855–867.
- Dunia, R. and Qin, S. J. (1998), ‘A unified geometric approach to process and sensor fault identification and reconstruction: The unidimensional fault case’, *Computers Chemical Engineering* **22**(7-8), 927–943.
- Duvenaud, D. K. (2014), Automatic Model Construction with Gaussian Processes, PhD thesis, University of Cambridge.
- Faes, L., Marinazzo, D. and Stramaglia, S. (2017), ‘Multiscale information decomposition: Exact computation for multivariate Gaussian processes’, *Entropy* **19**(8).
- Faes, L., Porta, A., Nollo, G. and Javorka, M. (2016), ‘Information Decomposition in Multivariate Systems: Definitions, Implementation and Application to Cardiovascular Networks’, *Entropy* **19**(1), 5.
- Farenzena, M. and Trierweiler, J. (2009), ‘LoopRank: a novel tool to evaluate loop connectivity’, *Proceedings of ADCHEM-2009* .
- Farenzena, M., Trierweiler, J. and Shah, S. (2009), ‘Variability matrix: A novel tool to prioritize loop maintenance’, *Proceedings of ADCHEM-2009* .
- Gao, B., Liu, T., Wei, W., Wang, T. and Li, H. (2011), Semi-supervised ranking on very large graphs with rich metadata, in ‘Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining’, number 49, pp. 96–104.

- Gao, D., Wu, C., Zhang, B. and Ma, X. (2010), ‘Signed Directed Graph and Qualitative Trend Analysis Based Fault Diagnosis in Chemical Industry’, *Chinese Journal of Chemical Engineering* **18**(2), 265–276.
- Hagberg, A. A., National, L. A. and Alamos, L. (2008), Exploring Network Structure, Dynamics, and Function using NetworkX, in ‘7th Python in Science Conference (SciPy 2008)’, pp. 11–15.
- Harel, D. (1988), ‘On visual formalisms’, *Communications of the ACM* **31**(5), 514–530.
- Hlaváčková-Schindler, K., Paluš, M., Vejmelka, M. and Bhattacharya, J. (2007), ‘Causality detection based on information-theoretic approaches in time series analysis’, *Physics Reports* **441**(1), 1–46.
- Iri, M., Aoki, K., O’Shima, E. and Matsuyama, H. (1979), ‘An algorithm for diagnosis of system failures in the chemical process’, *Computers & Chemical Engineering* **3**, 489–493.
- Jiang, H., Patwardhan, R. and Shah, S. L. (2009), ‘Root cause diagnosis of plant-wide oscillations using the concept of adjacency matrix’, *Journal of Process Control* **19**(8), 1347–1354.
- Kaiser, A. and Schreiber, T. (2002), ‘Information transfer in continuous processes’, **166**(March), 43–62.
- Kantz, H. and Schreiber, T. (2003), *Nonlinear Time Series Analysis*, Cambridge University Press.
- Kourti, T. and MacGregor, J. F. (1996), ‘Multivariate SPC methods for process and product monitoring’, *Journal of Quality Technology* **28**(4), 409–428.
- Kraskov, A., Stögbauer, H. and Grassberger, P. (2004), ‘Estimating mutual information’, *Physical Review E* **69**(6), 066138.
- Landman, R., Kortela, J., Sun, Q. and Jämsä-Jounela, S. L. (2014), ‘Fault propagation analysis of oscillations in control loops using data-driven causality and plant connectivity’, *Computers and Chemical Engineering* **71**, 446–456.
- Liu, J. and Chen, D.-S. (2014), ‘Fault isolation using modified contribution plots’, *Computers & Chemical Engineering* **61**, 9–19.
- Lizier, J. T. (2014), JIDT : An information-theoretic toolkit for studying the dynamics of complex systems, Technical report.

- Lizier, J. T., Heinzle, J., Horstmann, A., Haynes, J.-D. and Prokopenko, M. (2011), ‘Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity.’, *Journal of computational neuroscience* **30**(1), 85–107.
- Lizier, J. T., Prokopenko, M. and Zomaya, A. Y. (2008), ‘Local information transfer as a spatiotemporal filter for complex systems’, *Physical Review E* **77**(2), 026110.
- Lizier, J. T. and Rubinov, M. (2012), ‘Multivariate construction of effective computational networks from observational data’, *Max Planck Institute: Preprint* .
- Mader, W., Feess, D., Saur, D., Lange, R., Glauche, V., Weiller, C., Timmer, J. and Schelter, B. (2010), ‘Investigating Multivariate Systems using Directed Partial Correlation’, *Intern. J. Bioelectromagn.* **12**(1), 21–25.
- Marschinski, R. and Kantz, H. (2002), ‘Analysing the information flow between financial time series’, *The European Physical Journal B* **30**(2), 275–281.
- Maurya, M. (2003a), ‘A systematic framework for the development and analysis of signed digraphs for chemical processes. 1. Algorithms and analysis’, *Industrial & Engineering Chemistry Research* **17**, 4789–4810.
- Maurya, M. R. (2003b), Integrating causal models and trend analysis for process fault diagnosis, PhD thesis, Purdue University.
- Maurya, M. R., Rengaswamy, R. and Venkatasubramanian, V. (2003), ‘A Systematic Framework for the Development and Analysis of Signed Digraphs for Chemical Processes. 2. Control Loops and Flowsheet Analysis’, *Industrial & Engineering Chemistry Research* **42**(20), 4811–4827.
- Maurya, M. R., Rengaswamy, R. and Venkatasubramanian, V. (2006), ‘A signed directed graph-based systematic framework for steady-state malfunction diagnosis inside control loops’, *Chemical Engineering Science* **61**(6), 1790–1810.
- Maurya, M. R., Rengaswamy, R. and Venkatasubramanian, V. (2007), ‘A signed directed graph and qualitative trend analysis-based framework for incipient fault diagnosis’, *Chemical Engineering Research and Design* **85**(10 A), 1407–1422.
- Nemenman, I., Bialek, W. and de Ruyter van Steveninck, R. (2004), ‘Entropy and information in neural spike trains: Progress on the sampling problem’, *Physical Review E* **69**(5), 056111.
- Padmanabhan, D., Desikan, P., Srivastava, J. and Riaz, K. (2005), ‘WICER: A Weighted Inter-Cluster Edge Ranking for Clustered Graphs’, *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI’05)* pp. 522–528.

- Paluš, M., Komárek, V., Hrnčíř, Z. and Štěrbová, K. (2001), ‘Synchronization as adjustment of information rates: detection from bivariate time series.’, *Physical review. E, Statistical, nonlinear, and soft matter physics* **63**, 046211.
- Paulonis, M. and Cox, J. (2003), ‘A practical approach for large-scale controller performance assessment, diagnosis, and improvement’, *Journal of Process Control* **13**(2), 155–168.
- Pearl, J. (2009), *Causality: Models, Reasoning and Inference*, 2nd edn, Cambridge University Press, New York.
- Rahman, A. and Choudhury, M. S. (2011), ‘Detection of control loop interactions and prioritization of control loop maintenance’, *Control Engineering Practice* **19**(7), 723–731.
- Ricker, N. L. (1996), ‘Decentralized control of the Tennessee Eastman Challenge Process’, *Journal of Process Control* **6**(4), 205–221.
- Rodriguez, M. and Sanz, R. (2009), ‘Development of integrated functional-structural models’, *Computer aided chemical engineering* pp. 573–578.
- Rossi, M., Tangirala, A., Shah, S. and Scali, C. (2006), ‘A data-based measure for interactions in multivariate systems’, pp. 681–686.
- Runge, J., Heitzig, J., Petoukhov, V. and Kurths, J. (2012), ‘Escaping the curse of dimensionality in estimating multivariate transfer entropy’, *Physical Review Letters* **108**(25), 1–5.
- Schreiber, T. (2000), ‘Measuring Information Transfer’, *Physical Review Letters* **85**(2), 461–464.
- Schreiber, T. and Schmitz, A. (2000), ‘Surrogate time series’, *Physica D: Nonlinear Phenomena* **142**(3-4), 346–382.
- Shannon, C. E. (1948), ‘A Mathematical Theory of Communication’, *Bell System Technical Journal* **27**(3), 379–423.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003), ‘Cytoscape: a software environment for integrated models of biomolecular interaction networks.’, *Genome research* **13**(11), 2498–504.
- Shu, Y. and Zhao, J. (2013), ‘Data-driven causal inference based on a modified transfer entropy’, *Computers & Chemical Engineering* **57**, 173–180.

- Singh, A. and Lesica, N. a. (2010), ‘Incremental mutual information: a new method for characterizing the strength and dynamics of connections in neuronal circuits.’, *PLoS computational biology* **6**(12), e1001035.
- Skogestad, S. and Postlethwaite, I. (2005), *Multivariable Feedback Control*, Wiley.
- Staniek, M. and Lehnertz, K. (2008), ‘Symbolic transfer entropy’, *Physical Review Letters* **100**(15), 1–4.
- Steege, G. V. (2013), Non-parametric Entropy Estimation Toolbox (NPEET), Technical report, University of Southern California.
- Su, J., Wang, D., Zhang, Y., Yang, F., Zhao, Y. and Pang, X. (2017), ‘Capturing causality for fault diagnosis based on multi-valued alarm series using transfer entropy’, *Entropy* **19**(12), 5–8.
- Sugihara, G., May, R., Ye, H., Hsieh, C.-h. C.-h., Deyle, E., Fogarty, M. and Munch, S. (2012), ‘Detecting causality in complex ecosystems.’, *Science* **338**(6106), 496–500.
- Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007), ‘Measuring and testing dependence by correlation of distances’, *Annals of Statistics* **35**(6), 2769–2794.
- Tabaru, T., Nakano, K., Shin, S. and Matsuo, T. (2005), Plant data analysis by Gabor wavelet, in ‘Proceedings of the SICE Annual Conference’, Okayama, Japan, pp. 950–955.
- 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**(2), 503–512.
- Thornhill, N. F. and Horch, A. (2006), ‘Advances and new directions in plant-wide controller performance assessment’, *Proceedings of ADCHEM-2006* .
- Thornhill, N., Shah, S., Huang, B. and Vishnubhotla, A. (2002), ‘Spectral principal component analysis of dynamic process data’, *Control Engineering Practice* **10**(8), 833–846.
- Tsubakino, D. and Hara, S. (2012), ‘Eigenvector-based intergroup connection of low rank for hierarchical multi-agent dynamical systems’, *Systems & Control Letters* **61**(2), 354–361.
- Vakorin, V., Krakovska, O. and McIntosh, A. (2009), ‘Confounding effects of indirect connections on causality estimation’, *Journal of neuroscience . . .* **184**, 152–160.

- Venkatasubramanian, V., Rengaswamy, R. and Ka, S. N. (2003), ‘A review of process fault detection and diagnosis Part III : Process history based methods’, *Computers & Chemical Engineering* **27**(3).
- Vianna, R. F. and McGreavy, C. (1995), ‘Qualitative modelling of chemical processes - a weighted digraph (WDG) approach’, *Computers and Chemical Engineering* **19**(SUPPL. 1), 375–380.
- Weber, I., Florin, E., Von Papen, M. and Timmermann, L. (2017), ‘The influence of filtering and downsampling on the estimation of transfer entropy’, *PLoS ONE* **12**(11), e0188210.
- Weidl, G., Madsen, A. and Israelson, S. (2005), ‘Applications of object-oriented Bayesian networks for condition monitoring, root cause analysis and decision support on operation of complex continuous processes’, *Computers & Chemical Engineering* **29**(9), 1996–2009.
- Westerhuis, J. A., Gurden, S. P. and Smilde, A. K. (2000), ‘Generalized contribution plots in multivariate statistical process monitoring’, *Chemometrics and Intelligent Laboratory Systems* **51**(1), 95–114.
- Wibral, M., Vicente, R. and Lizier, J. (2014), *Directed Information Measures in Neuroscience*, Springer.
- Wilken, S. E. (2012), Ranking the Importance of Control Loops Using Google PageRank Inspired Eigenvector Analysis, Technical report, University of Pretoria, Pretoria.
- Yang, F., Duan, P., Shah, S. and Chen, T. (2014), *Capturing Connectivity and Causality in Complex Industrial Processes*, Springer.
- Yim, S., Ananthakumar, H., Benabbas, L., Horch, A., Drath, R. and Thornhill, N. (2006), ‘Using process topology in plant-wide control loop performance assessment’, *Computers & Chemical Engineering* **31**(2), 86–99.
- Yu, W. and Yang, F. (2015), ‘Detection of causality between process variables based on industrial alarm data using transfer entropy’, *Entropy* **17**(8), 5868–5887.
- Zalmijn, E. and Sigtermans, D. (2017), Identifying influential nodes in complex networks, in ‘Coupling and Causality in Complex Systems’, Cologne.
- Zhao, C., Sun, Y. and Gao, F. (2012), ‘A multiple-time-region (MTR)-based fault subspace decomposition and reconstruction modeling strategy for online fault diagnosis’, *Industrial & Engineering Chemistry Research* **51**, 11207–11217.

Zhukov, L. E. (2015), 'Centrality Measures Graph-theoretic measures Which vertices are important ?'.