Graph-based exergy FDI sensitivity analysis on the Tennessee Eastman process

E. Nortje

orcid.org/ 0000-0001-5848-7825

Dissertation accepted in fulfilment of the requirements for the degree *Master of Engineering in Electrical and Electronic Engineering* at the North-West University

Supervisor: Prof. K.R. Uren
Co-supervisor: Prof. G. van Schoor

Graduation: November 2024

Acknowledgements

I would like to thank Prof. K.R. Uren and Prof. G. van Schoor for their guidance throughout this study. Additional thanks to J. Vosloo for her assistance with the Tennessee Eastman process model used during this study and for providing additional material from her work with the TEP model.

I would like to thank my parents for their patience, support and guidance throughout this endeavour. Thank you for leading by example when demonstrating commitment to challenges accepted and perseverance to see them through. For this I am ever grateful.

Abstract

This study presents work on identifying the location and options possible for functional parameters associated with each energy graph-based visualisation (EGBV) method. Each of the EGBV methods were analysed to determine whether sensitivity to functional parameters exists during FDI of fault conditions using the Tennessee Eastman process (TEP) model. Additionally, validation of this sensitivity analysis was performed to confirm if initial results held true under different model operation modes. Literature was studied on existing Heterogeneous Euclidean-Overlap Metric (HEOM) based distance functions as well as previous implementations of the respective EGBV methods to identify possible alternative options for each functional parameter. Following the identification and compiling a structure of functional parameters for each EGBV method, the Tennessee Eastman process model was implemented to acquire data which demonstrate a range of various fault conditions during system operation. The structure utilised to acquire multiple datasets at the default base case mode later served useful in providing an opportunity to acquire datasets for validation purposes. Initial datasets that were acquired were utilised in transforming process measurements into the attributed graphs utilised for fault detection and isolation (FDI) of operational conditions observed. Each EGBV method was implemented as a range of functional parameter permutations in delivering FDI results that were quantitatively measured and analysed to identify possible sensitivity demonstrated.

The execution of the functional parameter permutations for each EGBV method delivered gave a clear indication of the impact demonstrated by functional parameter selection on each category of performance measured. A clear deviation in performance was observed when comparing alternate functional parameter options to the default methodology for each EGBV method. The sensitivity analysis further confirmed that sensitivity in performance was demonstrated towards the functional parameter locations investigated. The validation of the initial sensitivity analysis performed consisted in reconfiguring the TEP model to mode 5 operation, in presenting new conditions for sensitivity to be observed. This validation delivered both a confirmation of sensitivity and new information on how sensitivity demonstrated by the EGBV methods varied between model operation modes. This study presents evidence to the deviation in FDI performance resulting from EGBV method functional parameters and sensitivity in performance demonstrated towards each functional parameter. This sensitivity is validated for a single model application, and presents an initial case to investigate functional parameters at a larger scale in future research.

Keywords: Tennessee Eastman process, functional parameter, EGBV, HEOM, cost-matrix utilisation, sensitivity analysis, ANOVA

Contents

Abbreviations

EGBV	Energy graph-based visualisation
FDI	Fault Detection and Isolation
FP	False positive
HEOM	Heterogeneous Euclidean Overlap Metric
IQR-HEOM	Inter-quartile range Heterogeneous Euclidean Overlap Metric
MAD	Mean Absolute Deviation
NOC	Normal operating condition
SA	Sensitivity analysis
SHEOM	Scaled Heterogeneous Euclidean Overlap Metric
SD	Standard deviation
TEP	Tennessee Eastman Process
XMEAS	Process measure
XMV	Manipulated variable

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Fault detection and isolation techniques aim to reliably detect and determine the fault type, and location when isolating an abnormal condition status. Kurtoglu et al. defines fault definition as the binary classification, faulty or non-faulty, of the system state. The author also defines isolation as "the determination of the fault mode and location in the system [1]." FDI techniques have become a focal point for study with the goal of delivering actionable solutions to model systems as representation to industry conditions. The requirement for FDI techniques primarily stem from hardware redundancy. This can be model- or data-based systems where hardware redundancy-based deployment on physical systems is financially expensive to install and maintain, while it is also limiting the system to only identify known process conditions. Model-based FDI techniques demonstrate more flexible implementation during complex system applications where changes to process or structural changes could occur [2,3].

Model-based FDI further consists of sub-approaches where each utilise a different technique to deliver the classification of fault conditions observed. Knowledge-based models perform qualitative evaluation of observed values, founded on human knowledge-based rules when performing condition classification. Analytical methods perform a quantitative analysis of the presented model conditions whereas data-based approaches utilise recorded process data from a system or model in detecting and isolating fault conditions. Such a model aims to adapt to observed model conditions to best estimate output measurements affected by input parameters, instead of considering the model function when classifying observed conditions [3].

Energy graph-based visualisation (EGBV) methods incorporate attributed graphs based on structural analysis of the considered system, where directional graphs represent binary interactions. These interactions are arranged in a matrix format filled with the attribute values indicating quantitative interaction between respective system nodes [4]. Process measure-

ments delivered by both physical- and model-based systems consist of various measurements (current, temperature, pressure, flow, density, etc.) at fixed locations which do not readily contribute to node interactions, while both energy and exergy could be determined per node when derived from these process measures. This allows for calculating the change in exergy flow, using both incoming and outgoing exergy per node in tracking directional node interactions. This could then be considered a signature when observing the operating conditions of a respective process. The development in graph-based FDI techniques was presented in [4], where a Qualitative approach was used to evaluate the change in exergy flow patterns as signature during observed fault conditions for a heated two-tank system. Further investigation into the performance of EGBV methods was done by [5] and [6] where Qualitative and Quantitative approaches were implemented to perform condition classification. The FDI performance results presented by these authors point to possible sensitivity towards unknown sources of variation. These sources of variation to FDI performance could result from the functional components implemented within each of the FDI methods. These components are considered functional parameters where multiple options exist for each item (or location) to be implemented within the respective FDI methods. The selection of these functional parameters could deliver variation in FDI results when using the identical model data as input. This variation in performance could further indicate sensitivity demonstrated towards functional parameters implemented.

1.2 Problem statement

EGBV methods have been previously applied, but have not yet been evaluated in response to their sensitivity demonstrated to underlying functional parameters comprising these FDI methods. These functional parameters include alternate options that exist for the individual steps taken in the methodology of an EGBV method. Sensitivity demonstrated towards functional parameters would include an analysis of FDI performance as measured using classification accuracy, isolation and the Matthews Correlation Coefficient (MCC).

1.3 Research objectives

1.3.1 Identification of relevant EGBV functional parameters

Identification of prominent graph-based FDI methods requires surveying literature available to identify choices made by authors shaping their implementation of these methods. Such choices should be inspected to identify potential functional parameters and options available for these parameters that would influence the function of the respective FDI methods.

1.3.2 Data compilation

A structure is required where data acquisition would be executed to obtain multiple datasets using the TEP model. This includes the methodology used in executing model simulation and how process measurement data would be transformed to deliver attributed graphs to be used during FDI. Data compilation includes discussion on how the functional parameter combinations for each EGBV method would be implemented, accompanied by discussion of performance measures to be used for this study.

1.3.3 Sensitivity analysis of EGBV functional parameters

A sensitivity analysis of the identified EGBV methods and associated functional parameters would serve to identify the underlying sensitivity demonstrated. This would require performing FDI using the EGBV methods followed by using FDI results to perform the sensitivity analysis.

1.3.4 Validation of sensitivity analysis

Validation of the sensitivity analysis will be performed to establish whether the experiment's results meet the required level of accuracy in satisfying the problem statement identified for this study. This includes testing the validity of the sensitivity analysis concept under a different set of operating conditions or model mode as not yet experienced during this study. This would establish if the study concept is valid for broad or model specific use case only.

1.4 Methodology

1.4.1 Identification of relevant EGBV functional parameters

Existing literature would need to be surveyed to identify options available for selected functional parameters. These options include author decisions or additional developments made to similar methods that could further be implemented for the EGBV methods in this study.

1.4.2 Data compilation

The simulation model would be inspected to identify variables that could be used to control operating conditions in generating multiple datasets. Following the acquisition of datasets, a deliberation is presented on how this process data would be transformed into a usable format

by which FDI could be performed. Data compilation includes discussion of the performance measures by which FDI results were evaluated for the respective EGBV methods.

1.4.3 Sensitivity analysis of EGBV functional parameters

The EGBV functional parameter permutations were first executed using the datasets generated for this study. FDI results delivered by these EGBV permutations were used to perform the sensitivity analysis to determine whether any of the considered EGBV methods demonstrate sensitivity to the identified functional parameters.

1.4.4 Validation of sensitivity analysis

The TEP model would be configured to mode 5 operation to acquire new datasets, from which validation would be performed. The FDI and sensitivity analysis of the EGBV methods would be duplicated using these datasets to determine whether the observed results remain valid for this operating mode.

1.5 Dissertation Outline

Chapter 2 contains background knowledge on topics associated with the content of this study, followed by a critical review of literature which proved to be significantly important to this study. **Chapter 3** presents a discussion on the functional parameters identified and their respective options for each EGBV method. These functional parameters are then presented as a summary of permutations available for each EGBV method. **Chapter 4** provides information on the TEP model and its operation in simulating datasets for this study. The transformation of process measurement data into attributed graphs is discussed. In **Chapter 5** FDI is performed using the EGBV method permutations, followed by discussion on the FDI results observed. These FDI results are then used in performing sensitivity analysis of functional parameters. **Chapter 6** demonstrates the validation performed for FDI and sensitivity analysis by implementing the EGBV methods under alternate operating conditions by the TEP model. The results are then compared with the TEP operation modes implemented for this study. **Chapter 7** provides closure of this study by reflecting on the research objectives stated, and whether these objectives were achieved/successful.

Chapter 2

Literature study

2.1 Introduction

With this study investigating the functional parameters of EGBV methods and how sensitivity towards these functional parameters would be demonstrated, this chapter presents literature studied on key concepts identified in executing this investigation. These concepts include the Tennessee Eastman process, EGBV for fault detection & isolation, sensitivity analysis methods, and verification & validation. These concepts are briefly discussed in providing the initial background on how these concepts would tie into the steps followed for this study.

Literature sources providing critical knowledge on the respective EGBV methods to be used for this study are further discussed in-depth, as well as how decisions made in the application of these methods were demonstrated by the respective authors. Work presented by [5] and [6] were identified as critical literature sources, presenting recent applications of these EGBV methods where each of these authors elaborated on their choices made for the respective steps within these methods.

2.2 Tennessee Eastman process

The Tennessee Eastman process model was initially introduced by [7], with the authors presenting an accurate test problem demonstrating similar characteristics and restrictions associated with the physical industrial process, which it was based upon. The model seen in figure 2.1 was proposed in response to the growing interest for industrial models that could serve as benchmark for testing and verification in the development of industrial control theories [7]. Table 2.1 demonstrates the various reactant materials introduced to the system, along with the subsequent products delivered as output by the TEP. Gas B is an inert material that

requires being purged from the system to maintain optimal product output by the system.

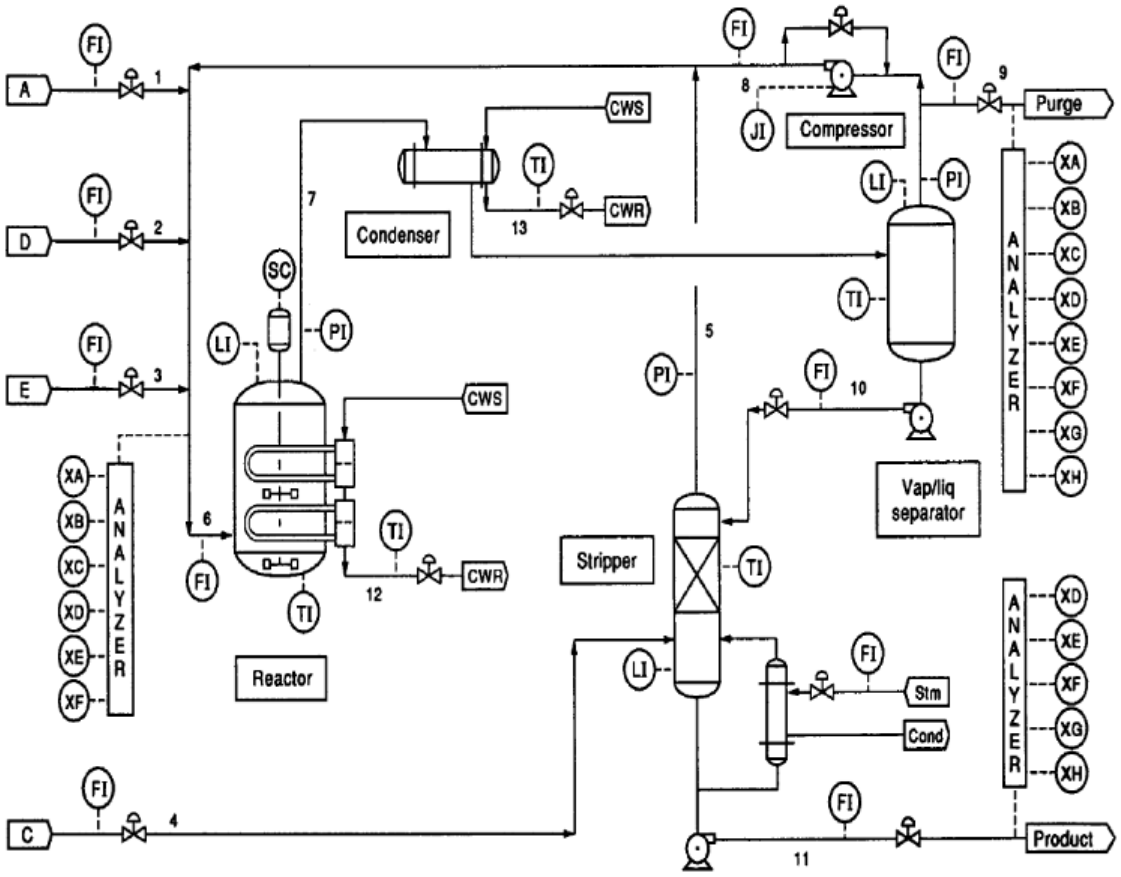

Figure 2.1: Tennessee Eastman test problem [7, Fig 1]

Table 2.1: System material components [7]

Component	**Phase**	**Type**
A	Gas	Reactant
B	Gas	Inert
C	Gas	Reactant
D	Gas	Reactant
E	Gas	Reactant
F	Liquid	Byproduct
G	Liquid	Product
H	Liquid	Product

Three gas-based reactants (A, D and E) are provided as input to the reactor, whereby an exothermic chemical reaction occurs aided by a non-volatile catalyst as liquid solution in the reactor, in order to deliver the two liquid products G and H as part of the output stream. Heat generated by the exothermic reaction is removed by means of an internal cooling bundle to the reactor. The reactor output consists of both gas and liquid components, and is passed through the condenser to reject the remaining heat from the mixed stream to allow further condensation of material vapour to liquid as required for the final product. This cooled material proceeds to the separator, where components in this stream are separated in a phase where condensed material is passed to the product stripper for further processing. During this process, non-condensed material is returned by the centrifugal compressor, looping back into the initial reactor input feed. Materials removed by this separator include byproduct F and the inert material created by the initial reaction process. Lastly, the condensed liquid mixture proceeds through the product stripper, where it is introduced to reactant C in removing the remaining initial reactant material still present in the mixture. At the same time, the remaining reactant material is returned via the recirculating loop connected as reactor input. The mixed product liquid consisting of products G and H are then delivered in the product stream exiting the TEP system [7]. Compositions of process products delivered by the TEP system are demonstrated in (1)-(4) [7].

$$A_{gas} + C_{gas} + D_{gas} = G_{liq} \tag{1}$$

$$A_{gas} + C_{gas} + E_{gas} = H_{liq} \tag{2}$$

$$A_{gas} + E_{gas} = F_{liq} \tag{3}$$

$$3D_{gas} = 2F_{liq} \tag{4}$$

Six operational modes have been outlined for the TEP model, seen in Table 2.2 below, where various ratios to product compositions would be delivered, depending on system constraints applicable, to maintain the profitability of the process. The initial TEP model by [7] implements an open-loop control system, which has since served as the platform for testing additional system control schemes, as this model contains 12 manipulated variables (XMV), accompanied by 41 measured variables (XMEAV) by which process conditions could be monitored. Tables 2.3 and 2.4 below list the respective variable types mentioned, with their number designations as published in [7].

Table 2.2: TEP system operation modes by Downs [7]

Mode	Ratio	Rate
1 (Base Case)	50/50	7038 kg/h G and 7038 kg/h H
2	10/90	1408 kg/h G and 12,669 kg/h H
3	90/10	10,000 kg/h G and 1111 kg/h H
4	50/50	maximum production rate
5	10/90	maximum production rate
6	90/10	maximum production rate

Table 2.3: Process manipulated variables [7, Tab. 3]

Variable	Name	Low Limit	High Limit	Unit
XMV(1)	D feed flow	0	5811	kg/h
XMV(2)	E feed flow	0	8354	kg/h
XMV(3)	A feed flow	0	1.017	kscm/h
XMV(4)	A and C feed flow	0	15.25	kscm/h
XMV(5)	Compressor recycle valve (%)	0	100	-
XMV(6)	Purge valve (%)	0	100	-
XMV(7)	Separator pot liquid flow	0	65.71	m^3/h
XMV(8)	Stripper liquid product flow	0	49.10	m^3/h
XMV(9)	Stripper steam valve (%)	0	100	-
XMV(10)	Reactor cooling water flow	0	227.1	m^3/h
XMV(11)	Condenser cooling water flow	0	272.6	m^3/h
XMV(12)	Agitator speed	150	250	rpm

Process measurements defined in Table 2.4 comprise of multiple units of measure, in providing an opportunity for detailed monitoring of the process operation while additionally increasing the difficulty associated with the incorporation of all unit dimensions, depending on the respective use case. The remaining process measurements solely consist of material compositions per the following locations in the TEP model:

- Reactor feed analysis of materials A to F are XMEAS(23) to XMEAS(28) respectively.
- Purge gas analysis of materials A to H are XMEAS(29) to XMEAS(36) respectively.
- Product analysis of materials D, E, F, G and H are XMEAS(37) to XMEAS(41) respectively.

Table 2.4: Continuous process variables [7, Tab. 4]

Variable	**Name**	**Unit**
XMEAS(1)	A Feed	kscm/h
XMEAS(2)	D Feed	kscm/h
XMEAS(3)	E Feed	m^3/h
XMEAS(4)	A and C Feed	m^3/h
XMEAS(5)	Reactor flow	kscm/h
XMEAS(6)	Reactor feed rate	kscm/h
XMEAS(7)	Reactor pressure	kPa
XMEAS(8)	Reactor level (%)	-
XMEAS(9)	Reactor temperature	°C
XMEAS(10)	Purge rate	kscm/h
XMEAS(11)	Product separator temperature (%)	-
XMEAS(12)	Product separator level	°C
XMEAS(13)	Product separator pressure	kPa
XMEAS(14)	Product separator underflow	m^3/h
XMEAS(15)	Stripper level (%)	-
XMEAS(16)	Stripper pressure	kPa
XMEAS(17)	Stripper underflow	m^3/h
XMEAS(18)	Stripper temperature	°C
XMEAS(19)	Stripper steam flow	m^3/h
XMEAS(20)	Compressor work	kW
XMEAS(21)	Reactor cooling water outlet temperature	°C
XMEAS(22)	Separator cooling water outlet temperature	°C

The TEP model includes 21 process faults that could be introduced during simulation of the system and consist of 5 different types based on how these faults present during operation. These faults are presented in Table 2.5 including details to the process variable which is affected during each of these faults. Faults 1-7 occur as a step-change in the associated process variable while faults 8-12 present as an increased variability of some process variables [7, 8]. Fault 21 is associated with a sticking valve where the valve for process stream 4 is fixed at a steady state position and was not implemented during this study.

Table 2.5: Process faults [8, Tab. 8.4]

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

Since its release by Downs and Vogel [7], the TEP model has been used as a benchmark to develop additional system control schemes, operating modes, or optimisation configurations as expansion upon the original TEP presented. Notable literature sources were identified, where authors present novel solutions in improving the system control and performance delivered as further development of the initial model by Downs [7]. A brief summary of notable control strategies developed shortly after publication of the initial TEP model is presented below:

Ricker

Ricker presents an optimised solution in satisfying operating constraints, while maintaining optimal operation during the steady-state operation of the model, by monitoring multiple inputs simultaneously. This includes addressing the minimum cost and maximum production problem functions separately to find the best suited operation point during system operation. Ricker concludes that this scheme lacks robustness in providing a solution for all operating modes considered, and should be researched further before considered suitable for application [9].

McAvoy and Ye

McAvoy and Ye present their approach to developing a plant system controller for the TEP model, based on a series of stages by which control is distributed. Their work consists of closed controller feedback loops to form the first layer in a cascaded controller hierarchy, and contributes in reducing the impact of disturbances on system operation. Their control scheme additionally includes closing the analyser loops and perceiving process stream compositions as measured variables. The authors then performed a relative gain analysis to determine loop pairings and evaluate their stability when compared to the Niederlinski Index. Lastly, the controllers arranged in cascaded hierarchy levels were tuned to deliver the best disturbance response in order to deliver optimal system operation [10].

Lyman and Georgakis

These authors presented four control schemes for plant-wide control of the TEP model, utilizing single-input single-output controllers in mitigating process disturbances experienced during system operation. Their work presents an in-depth discussion on the controller layouts as either standalone or cascaded, and how their functionality maintains a stable operation of the closed loop system. What is noteworthy is that one of these control structures is capable in withstanding all process disturbances without reaching the shutdown point for the system, and was recommended as a successful solution to the plant-wide control problem for this process model [11].

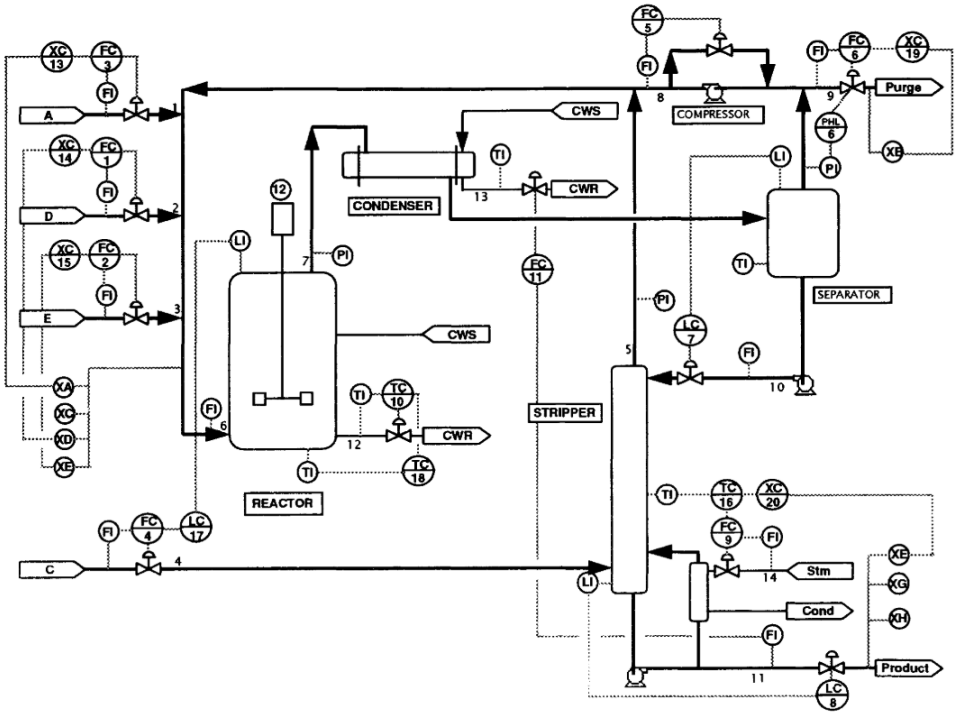

Figure 2.2: TEP Process - control structure 2 [11, Fig. 1]

The control schemes presented by these authors have since served as a foundation in literature, where the shift in focus has transitioned towards implementing the TEP model during the development of fault detection, isolation and/or diagnosis methods. The TEP model has, since its initial release, also been adapted towards educational purposes as a training tool and case study in industrial control, as done by [8], where the TEP model was demonstrated using one of the control schemes developed by [11]. Chiang et al. [8] then presents an in-depth discussion on the simulation conditions, datasets obtained during operation, and performance demonstrated by the respective fault detection method utilised [8]. A recent development utilising the Tennessee Eastman process model by [12] includes recreating the TEP simulation using a MATLAB software package in delivering a modernised version of the simulation model. Vosloo's [12] instance of the TEP model provides clear access to all main process subsystems at the component level, including the system controllers responsible for monitoring and manipulating system parameters.

2.3 Fault detection

A fault is defined in [13] as the unpermitted deviation of a system parameter or property beyond boundaries defined for safe operational conditions. The occurrence of a fault condition can further be described according to an individual component, form, and how the fault contributes to the system:

1. A component fault occurs during failure of one or more sub-components in the system at either hardware or software level in a system. Such faults would result in incorrect process measurements, impeding equipment function, false alarms, or process failure in the system [13, 14].

2. The form of a fault indicates to the rate of change and cumulative duration of the respective deviation observed. Fault deviation may present continuously in process measurements observed or in an abrupt manner, as seen with equipment malfunction when transitioning into a system fault condition. The abrupt fault form presents as a step fault condition, and is commonly the fastest deviation amidst the considered condition types. The incipient fault form occurs as a gradual drift in process conditions, which would only be discernible at a later time or would be omitted if the rate of change was faster than the defined sampling rate. The intermittent fault form presents as repeated switching between active and inactive states without remaining as an active fault state, and could lead to the possibility in perceiving the fault as a self-correcting condition [13, 14].

3. Lastly, the manner by which a fault contributes to the system state is of note, as the fault condition could be:
   - Additive as demonstrated by constant linear deviation or offset bias present in process measurements.
   - Multiplicative where variation to faults include surges or stuttering to be observed in process conditions [13, 14].

2.4 Fault isolation

Following fault detection, the fault would then need to be further evaluated to determine associated properties [13]:

- Location
- Kind
- Time of detection
- Size
- Time-variant behaviour

One of the following functions could be implemented depending on the scope of fault properties required:

Fault isolation: Determine fault kind, location and time of detection.
Fault identification: Follows fault isolation and includes fault size and time-variant behaviour.
Fault diagnosis: Determine fault kind, size, location and time of detection [6, 13].

Focusing on fault isolation for this study, a fault can only be isolated if the fault is unique to a specific case and distinction could be made from different faults that could occur for the system [6]. This would include discerning the operational condition, as either normal or fault condition, from the range of reference conditions implemented for this study.

2.5 Graph-based FDI

Graph-based fault detection and isolation is based on comparing attributed node graphs, representing a process system [4], in order to correctly discern the observed operation condition from the ground truth reference condition as either normal or fault condition. Jouili [15] performed graph matching using object images, where matching according to similarity using the Heterogeneous Euclidean Overlap Metric (HEOM) to determine node-to-node graph distances as measure in similarity [15]. Each attributed graph, example seen in Figure 2.3, consists of nodes interconnected by edges with each node containing an attribute. These attributes could be numerical and/or symbolical in representing the node signatures used during the graph matching process to determine similarity graphs.

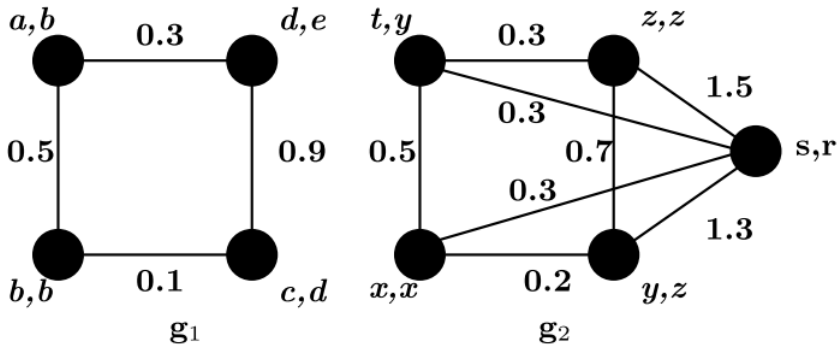

Figure 2.3: Two attributed graphs [15, Fig. 1]

Van Graan et al. [16] presents fault diagnosis of a counter-flow heat exchanger through energy-based graph matching. Energy characteristics were calculated for the heat exchanger system using data gathered during normal operation and three different fault conditions experienced during system operation to allow creating attributed graphs for each of the conditions observed. Each of the fault operation graphs used for the comparison where matched against the normal operation graph to determine the distance between graphs. Distance between corresponding nodes were calculated and arranged in a cost matrix were eigendecomposition was performed to deliver eigenvalues as a means to uniquely identify each condition. Figure 2.4 presents the signatures, containing both real- and imaginary signatures, for each of the conditions observed and the normal condition used as reference for case 2 in verifying if the three faults were identifiable [16].

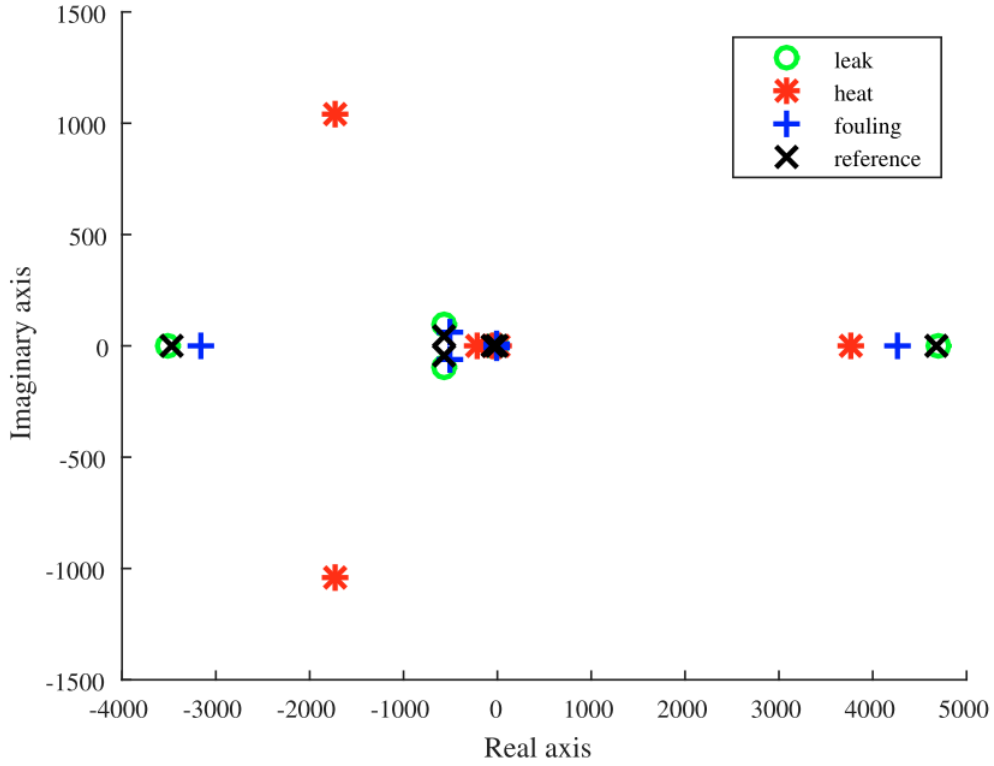

Figure 2.4: Case 2: Reference signature and energy visualisations of faults [16, Fig. 8]

Uren et al. [4] performed energy attributed graph-based FDI as applied to a two-tank system, where focus was on implementing a node signature matrix consisting of both change in exergy flow rate and energy flow rate. Following graph matching and calculating the cost matrix, eigendecomposition was performed on the cost matrix where the resulting eigenvalues were qualitatively assigned. These qualitative values were then arranged as identifier sets by which graphs were evaluated during fault detection and isolation. This approach was able to discern between normal- and fault conditions respectively when evaluating the qualitative difference signature compared to the signature of the respective reference conditions. This graph-based approach was then demonstrated as three different methods by [5] by performing FDI using a steady-state system model while [6] performed FDI using historical operational data , seen in figure 2.4 below, where each operation condition could be represented by a series of attributed graphs when performing FDI.

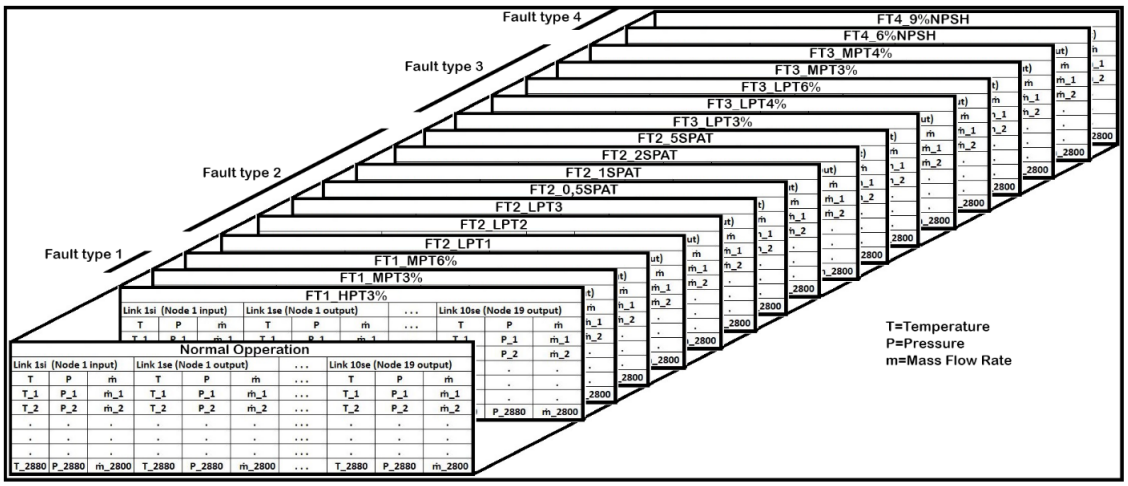

Figure 2.5: Illustration of the final time series data obtained for each of the simulated fault states and the normal operating state [6, Fig. 5.10]

2.6 FDI performance measures

The analysis of any FDI approach requires the means by which performance could be determined in executing its detection and isolation functionality. Quantitative performance metrics demonstrated in [17] and [1] include detection- and isolation metrics , sensitivity and stability factor for fault conditions experienced by the respective systems. Such characteristics include the following:

- The approach demonstrates the ability to quickly deliver a classification as whether a normal or fault condition is being observed. Any sensitivity demonstrated to noise could result in frequent misclassifications [18].
- The ability to correctly distinguish between known fault conditions is a trade-off in choice, between high isolability and accommodating model or process operation uncertainties [18].
- Robustness to noise and process uncertainties would prevent significant performance degradation from occurring when presented with varying process conditions [18].

Quantitative performance metrics demonstrated in [17] and [1] include detection and isolation metrics accompanied by the sensitivity and stability factor for fault conditions that are experienced by the respective systems. These factors determine an algorithm's sensitivity towards fault size at which detection is possible, as well as its stability in maintaining the correct detection classification during the occurrence of the fault condition [1]. The condition classifications presented below reflect fault isolation per the context of this study, as adapted from [17] and [19].

Table 2.6: FDI condition classification.

Detection	
False positive (FP)	Algorithm observes a fault condition in contrast to the ground truth being a normal operating condition.
False negative (FN)	Algorithm observes normal operation while the ground truth is a known fault condition.
True positive (TP)	Algorithm and ground truth are in alignment to the presence of a fault condition being observed.
True negative (TN)	Algorithm and ground truth are in alignment to the presence of a normal operating condition being observed.
Isolation	
Correct isolation	Condition detected and individually identified, corresponds with the ground truth condition.
Missed isolation	Condition detected and individually identified, does not correspond with the ground truth condition.

These classifications could then be compiled as a set for the range of operation conditions associated with a respective model to determine the quantitative rate for each classification [1].

Frank [20] presents a survey on metrics and methods available in assessing fault detection and diagnosis performance, where the Matthews Correlation Coefficient (MCC) is illustrated as an opportunity to include the true negative classification when calculating the detection rate performance, as demonstrated in (5) below [20]. The utilisation of the MCC is ideal as an additional measure to asses FDI methods, as it considers the TN rate in equal importance to the TP rate. Omitting detection of the NOC when performing FDI would cast doubt on the credibility of the true positive detection rate delivered by an FDI approach, when other measures do not provision for the NOC to be uniquely discerned for a considered system [20].

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{5}$$

2.7 Sensitivity analysis

Razavi et al. states "Sensitivity analysis (SA), in the most general sense, is the study of how the 'outputs' of a 'system' are related to, and are influenced by, its 'inputs'." Razavi further elaborates that "Inputs of interest, commonly referred to as 'factors' in SA, may include model parameters, forcing variables, boundary and initial conditions, choices of model structural configurations, assumptions and constraints." [21] Kleijnen presents an initial methodology in executing a sensitivity analysis, by considering the simulation model as a "black box" by which internal functionality was considered unknown with only input-output factors available for interaction [22]. A regression analysis could be performed to determine the main and interaction effects presented by factors. Fasso [23] presents a discussion on how the nature of the model utilised determines which sensitivity analysis method would be applicable for use.

Computer model

Computer models are used to represent complex systems, either linear or non-linear, implementing large quantities of input parameters where significance of parameter influence on the system output is not necessarily known. Sensitivity analysis of such systems would aid in reducing model complexity by removing input parameters with low influence on model system output. Analysis results could further be utilised towards model or concept problem validation [23].

Statistical model

The sensitivity analysis of statistical models is useful in discerning the level of dependence demonstrated by model results in regard to the stated hypotheses. This sensitivity could further be influenced by the nature of data or techniques implemented during model execution. Changes introduced to the input of statistical models could additionally indicate the presence of influential subsets within the observed data, leading to the possible occurrence of gross modelling errors resulting from deviations caused by these subsets. The identification of influential observations would additionally indicate the suitability of the considered model in addressing the respective study problems [21–23].

Regression model

An example of a regression model considered during sensitivity analysis would be an objective function, either linear or non-linear, where significant influence could occur in the event of outlier parameter values being observed during the evaluation of the respective function. This analysis would be applied in calculating the localised sensitivity demonstrated by the surrounding area to the optimal solution area for the considered function [23, 24].

Time series

An analysis of time series data presents the case where, constant change to multiple input factors could be experienced during the modelling of a system throughout the operation. An example of this system is used during environmental statistics, where meteorological data containing numerous factors is used for monitoring and forecasting purposes. Challenges are associated with the analysis of time series data, where a model could present sensitivity to outliers present within data, or the dataset considered is of low quality [23].

Sensitivity analysis approaches commonly used when considering the analysis of computer model systems are local sensitivity analysis (LSA) and global sensitivity analysis (GSA). LSA is performed by varying a single parameter at a time (one-at-a-time), while maintaining all other considered inputs at their predetermined baseline values to determine significant parameter influence on the model outcome. Alternatively, GSA consists of evaluating multiple system operation parameters, where parameters are varied simultaneously to deliver parameter combinations. This variation to output parameter values is then analysed per the respective input parameter variation, to calculate the significance of multiple order factor sensitivity indices with or without interactions, depending on the number of parameters utilised. The computational cost in executing the respective model must be considered when deciding which SA approach to implement, as high execution cost would become a limiting factor when performing analysis of a high number of input parameters [21, 23, 25].

2.8 Verification and validation

Verification and validation serve to establish the correctness of a proposed concept or model, and whether the considered concept or model is fit for its intended use. By performing these actions, the credibility in use of the concept or model could be established [26].

Different stages of validation are available for use, where conceptual validation consists of evaluating whether the theories and assumptions associated with a concept or model are correct and reasonable. This presents as the first phase, where factor screening should be performed to establish its significance to the considered concept. This should be achieved initially during the concept development, where operational validation would be performed closer to the end of the development cycle. Operational validation would determine whether the end result would be the observed outcome or the expected outcome. This will establish if any errors or deficiencies exist in the development of the respective concept or model [22, 26].

2.9 Critical literature review

Literature surveyed on EGBV methods present definitive development in the methodology to the concept where recent development by [5] and [6] presented the most significant contribution. Greyling presented a compelling discussion on steps followed in the EGBV methodology, which served as the starting point from which functional parameter location identification was performed in this study. Smith performed FDI using a series of attributed graphs to represent operational conditions observed. This delivered FDI classifications proportionate to the share of total operational graphs, demonstrating the highest similarity to each respective reference graph conditions. This methodology was identified to be of critical importance in performing similar FDI using the TEP model for time series operational conditions [6].

Following the initial identification of function parameter locations, literature which proved to be critical in the identification of possible alternatives suitable for use in the EGBV methods was by [27], [28] and [29]. Wilson & Martinez [27] presented a comprehensive survey performed on distance functions available, which were identified as a limitation in this study, followed by a discussion on the role of normalisation during the execution of the Heterogeneous Euclidean Overlap Metric (HEOM). Spencer and Dalatu [28, 29] then each presented their development on the initial HEOM to deliver the SHEOM and IQR-HEOM respectively as alternate normalisation options.

Sargent [26] presents an invaluable discussion on validation and verification of model or concept. This includes concept validation, which was identified to be a necessary step to be performed after identifying functional parameters for each EGBV method in this study. Sargent additionally reinforces the thought on considering the concept being validated to be fit for its intended purpose, and would serve as a guiding tool for the validation performed in this study. Lastly, [30] proved to be a significant literature source for the methodology presented in performing two-way ANOVA, as would be applied in this study.

Chapter 3

EGBV functional parameter identification

3.1 Introduction

This chapter presents a discussion on identified energy graph-based visualisation (EGBV) methods applied towards fault detection and isolation of fault conditions (FDI) observed during the operation of a dynamic system model. The researcher investigated underlying functional parameters identified from existing literature, focusing on locations where alternative options to underlying methodology could be implemented. Lastly, the implementation of the respective functional parameter permutations per FDI method are presented to best illustrate the difference between these permutations [31].

3.2 Energy-graph based FDI

This section proceeds with inspecting the components comprising the configuration of the EGBV methods (see Figure 3.1 below), where the process starts with obtaining operational data delivered by a system. This system data is first transformed by calculating the respective exergy values during system operation as a means of adding value to the observed data. The following step consists of transforming the TEP model data into an attributed graph arrangement according to the node signature matrix, which is then filled with system exergy data per operational data sample. The proposed node signature matrix is used for constructing the operation- and reference graphs when performing the Heterogeneous Euclidean Overlap Metric (HEOM) in calculating the graph distance between two graphs [15, 27].

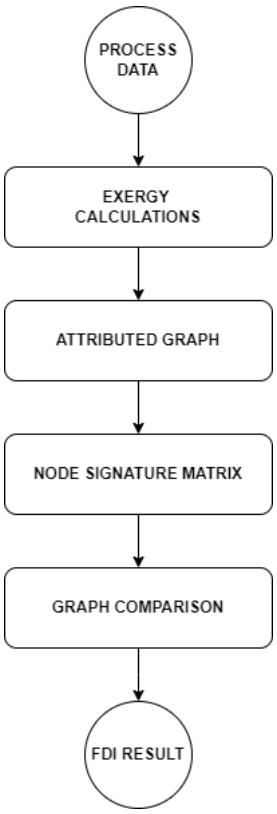

Figure 3.1: Graph-based FDI overview

3.2.1 Node signature matrix

The TEP model utilised during this study allows for exergy to be measured per individual system node where notation during this study follows on the model presented by [12], as illustrated in Figure 3.2. The system is divided into 15 nodes, including the external environment as 15th node in the system model. Both energy and exergy values were computed using calculations provided as extension to the TEP model developed in MATLAB by [12]. Previous work by [12] includes developing the necessary energy- and exergy calculations for the TEP system model and utilising these calculations to successfully perform exergy-based FDI using this system. These calculations require in-depth understanding of thermodynamic processes associated with chemical material processing in order to be verified. This falls beyond the scope of this study and would not contribute new information when attempted. Therefore these energy- and exergy calculations are assumed to be correct based on the successful results demonstrated by [12].

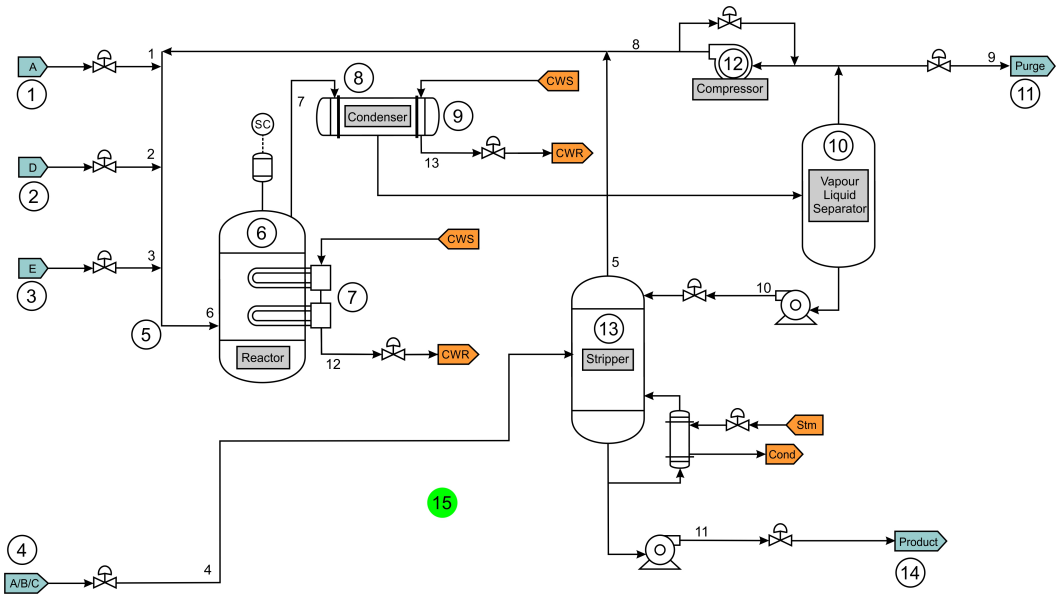

Figure 3.2: TEP system with graph notations provided by [12]

3.2.2 Attributed graph

A complex system with multiple sub-components can be defined as an attributed graph denoted $G = (N, L, A)$, with a defined set of non-empty system nodes N with links L, synonymous with edges, defining interconnections between respective nodes. Definition of the attribute set $A = a_{n_i}, a_{l_{ij}}$ in [12], with a_{n_i} defining the attribute of node a_{n_i} and $a_{l_{ij}}$ demonstrating the link attribute as from n_i to n_j with directional notation [4]:

- Edges from nodes were denoted as **positive**.
- Edges towards nodes were denoted as **negative**.

Following the node convention by [12] as illustrated in Figure 3.2 displaying the TEP model, the node notation representation in Figure 3.3 displays the edge interactions between system nodes as graph attributes. Note that node 15 represents the reference environment with which exergy calculations were performed. These graph attributes are then packaged using a similar node signature matrix defined by [4], where the node signature matrix was expanded to include two additional node attribute columns as first entries of the matrix to include the two exergy types, physical and chemical, which are present in the TEP system. The node signature matrix with (row, column) expanded from $(n \times n)$ to $(n \times (n+2))$, with (6) depicting the general node signature matrix with size (15×17) [4, 15].

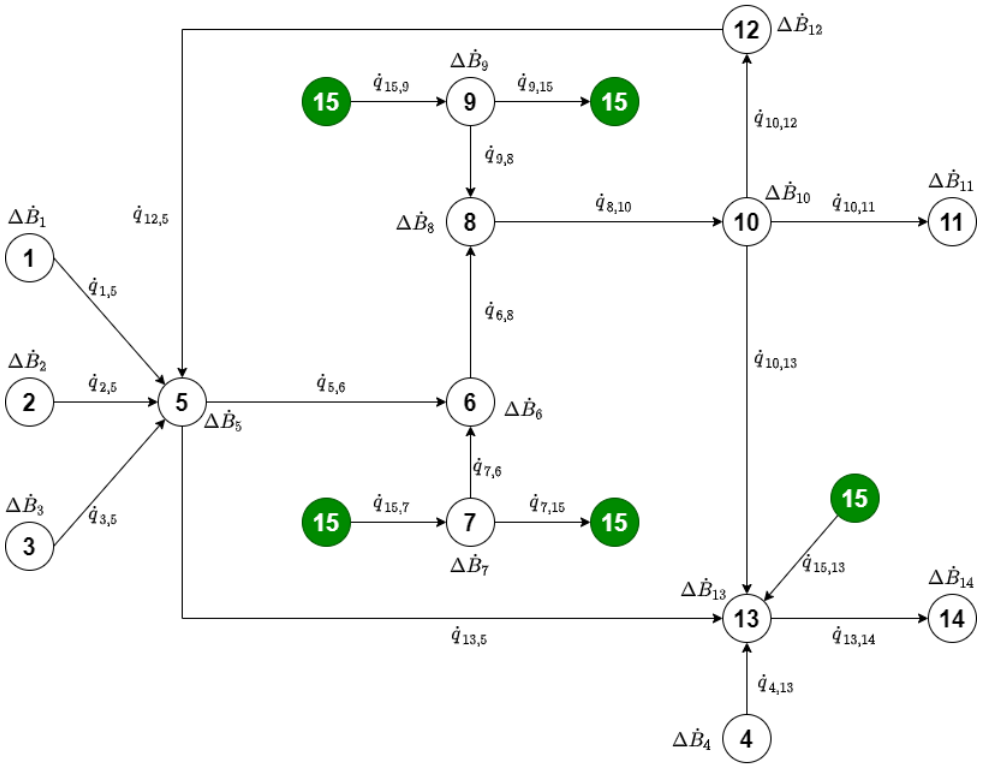

Figure 3.3: TEP system as graph structure by [12]

$$\mathbf{N_s} = \begin{bmatrix} \Delta\dot{B}_{ph1} & \Delta\dot{B}_{ch1} & \dot{Q}_{1,1} & \dot{Q}_{1,2} & \dot{Q}_{1,3} & \dot{Q}_{1,4} & \dot{Q}_{1,5} & \dot{Q}_{1,6} & \dot{Q}_{1,7} & \cdots & \dot{Q}_{1,15} \\ \Delta\dot{B}_{ph2} & \Delta\dot{B}_{ch2} & \dot{Q}_{2,1} & \dot{Q}_{2,2} & \dot{Q}_{2,3} & \dot{Q}_{2,4} & \dot{Q}_{2,5} & \dot{Q}_{2,6} & \dot{Q}_{2,7} & \cdots & \dot{Q}_{2,15} \\ \Delta\dot{B}_{ph3} & \Delta\dot{B}_{ch3} & \dot{Q}_{3,1} & \dot{Q}_{3,2} & \dot{Q}_{3,3} & \dot{Q}_{3,4} & \dot{Q}_{3,5} & \dot{Q}_{3,6} & \dot{Q}_{3,7} & \cdots & \dot{Q}_{3,15} \\ \Delta\dot{B}_{ph4} & \Delta\dot{B}_{ch4} & \dot{Q}_{4,1} & \dot{Q}_{4,2} & \dot{Q}_{4,3} & \dot{Q}_{4,4} & \dot{Q}_{4,5} & \dot{Q}_{4,6} & \dot{Q}_{4,7} & \cdots & \dot{Q}_{4,15} \\ \Delta\dot{B}_{ph5} & \Delta\dot{B}_{ch5} & \dot{Q}_{5,1} & \dot{Q}_{5,2} & \dot{Q}_{5,3} & \dot{Q}_{5,4} & \dot{Q}_{5,5} & \dot{Q}_{5,6} & \dot{Q}_{5,7} & \cdots & \dot{Q}_{5,15} \\ \Delta\dot{B}_{ph6} & \Delta\dot{B}_{ch6} & \dot{Q}_{6,1} & \dot{Q}_{6,2} & \dot{Q}_{6,3} & \dot{Q}_{6,4} & \dot{Q}_{6,5} & \dot{Q}_{6,6} & \dot{Q}_{6,7} & \cdots & \dot{Q}_{6,15} \\ \Delta\dot{B}_{ph7} & \Delta\dot{B}_{ch7} & \dot{Q}_{7,1} & \dot{Q}_{7,2} & \dot{Q}_{7,3} & \dot{Q}_{7,4} & \dot{Q}_{7,5} & \dot{Q}_{7,6} & \dot{Q}_{7,7} & \cdots & \dot{Q}_{7,15} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \Delta\dot{B}_{ph15} & \Delta\dot{B}_{ch15} & \dot{Q}_{15,1} & \dot{Q}_{15,2} & \dot{Q}_{15,3} & \dot{Q}_{15,4} & \dot{Q}_{15,5} & \dot{Q}_{15,6} & \dot{Q}_{15,7} & \cdots & \dot{Q}_{15,15} \end{bmatrix} \tag{6}$$

Substituting existing node edges from Figure 3.3 [12] into the matrix from (6) results in entries with zero values during (7) for non-existing edge connections between nodes, including the off-centre matrix diagonal $(n, n+2)$ which indicates self-referencing node edges. This node signature matrix is implemented when performing graph distance calculations during the graph-matching methods discussed in the next section.

$$\mathbf{N_s} = \begin{bmatrix} \Delta\dot{B}_{ph1} & \Delta\dot{B}_{ch1} & 0 & 0 & 0 & 0 & +\dot{Q}_{1,5} & 0 & 0 & \cdots & 0 \\ \Delta\dot{B}_{ph2} & \Delta\dot{B}_{ch2} & 0 & 0 & 0 & 0 & +\dot{Q}_{2,5} & 0 & 0 & \cdots & 0 \\ \Delta\dot{B}_{ph3} & \Delta\dot{B}_{ch3} & 0 & 0 & 0 & 0 & +\dot{Q}_{3,5} & 0 & 0 & \cdots & 0 \\ \Delta\dot{B}_{ph4} & \Delta\dot{B}_{ch4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ \Delta\dot{B}_{ph5} & \Delta\dot{B}_{ch5} & -\dot{Q}_{5,1} & -\dot{Q}_{5,2} & -\dot{Q}_{5,3} & 0 & 0 & +\dot{Q}_{5,6} & 0 & \cdots & 0 \\ \Delta\dot{B}_{ph6} & \Delta\dot{B}_{ch6} & 0 & 0 & 0 & 0 & -\dot{Q}_{6,5} & 0 & -\dot{Q}_{6,7} & \cdots & 0 \\ \Delta\dot{B}_{ph7} & \Delta\dot{B}_{ch7} & 0 & 0 & 0 & 0 & 0 & -\dot{Q}_{7,6} & 0 & \cdots & +\dot{Q}_{7,15} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \Delta\dot{B}_{ph15} & \Delta\dot{B}_{ch15} & 0 & 0 & 0 & 0 & 0 & 0 & +\dot{Q}_{15,7} & \cdots & 0 \end{bmatrix} \tag{7}$$

3.2.3 Operational graph

The operational graph represents the system operational data as an unknown condition prior to implementation of the FDI EGBV method. This will be compared with the set of known condition reference graphs during each FDI method in classifying the condition observed. A series of operational graphs would constitute the period of system operation observed for a respective operation condition of a system or model. Considering the position s of a graph in the overall series S of graphs, notation $G_{OP(condition)}(s)$ is utilised to represent the operation graph in position s for a respective condition. Using this series of operational graphs S during graph matching presents an opportunity in performing FDI with the ability to discern graph similarity over a complete operation period [5, 6].

3.2.4 Reference graph

Each known operation condition associated with the TEP model used during this study will require a point of reference serving as known truth, to which graph similarity will be measured when performing graph-based fault detection and isolation. Each operation's condition reference graph will consist of first calculating the average node value for each entry in the defined attribute graph matrix from (6), using a separate set of operational data than data used during the operational graphs above, in establishing expected values associated with a known condition instance. This reference graph per operation condition would then serve as the known truth to which similarity will be measured during graph comparison. Equation (8) demonstrates how the mean value per node position is calculated using the sum of values of each attributed graph instance s in series S.

$$Avg(i,j) = \frac{\sum_{s=1}^{S} N_s(i,j)}{S} \tag{8}$$

Following this notation, an example of the compiled reference graph for an operational condition is demonstrated below in (9), with $G_{REF(Condition)}$ as notation of the respective condition, followed by the attributed matrix from (6), populated with the average node value using (8).

$$\mathbf{G_{REF(condition)}} = \begin{bmatrix} Avg_{1,1} & Avg_{1,2} & Avg_{1,3} & Avg_{1,4} & Avg_{1,5} & Avg_{1,6} & Avg_{1,7} & \cdots & Avg_{1,17} \\ Avg_{2,1} & Avg_{2,2} & Avg_{2,3} & Avg_{2,4} & Avg_{2,5} & Avg_{2,6} & Avg_{2,7} & \cdots & Avg_{2,17} \\ Avg_{3,1} & Avg_{3,2} & Avg_{3,3} & Avg_{3,4} & Avg_{3,5} & Avg_{3,6} & Avg_{3,7} & \cdots & Avg_{3,17} \\ Avg_{4,1} & Avg_{4,2} & Avg_{4,3} & Avg_{4,4} & Avg_{4,5} & Avg_{4,6} & Avg_{4,7} & \cdots & Avg_{4,17} \\ Avg_{5,1} & Avg_{5,2} & Avg_{5,3} & Avg_{5,4} & Avg_{5,5} & Avg_{5,6} & Avg_{5,7} & \cdots & Avg_{5,17} \\ Avg_{6,1} & Avg_{6,2} & Avg_{6,3} & Avg_{6,4} & Avg_{6,5} & Avg_{6,6} & Avg_{6,7} & \cdots & Avg_{6,17} \\ Avg_{7,1} & Avg_{7,2} & Avg_{7,3} & Avg_{7,4} & Avg_{7,5} & Avg_{7,6} & Avg_{7,7} & \cdots & Avg_{7,17} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ Avg_{15,1} & Avg_{15,2} & Avg_{15,3} & Avg_{15,4} & Avg_{15,5} & Avg_{15,6} & Avg_{15,7} & \cdots & Avg_{15,17} \end{bmatrix} \tag{9}$$

The list of reference graphs compiled for all TEP model operation conditions considered during this study, as NOC and 20 fault conditions mentioned during Chapter 2, will constitute the reference graph database as illustrated in Figure 3.4. It will use the above-mentioned notation to indicate the respective operation condition observed.

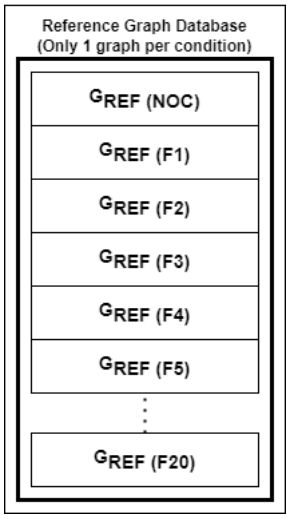

Figure 3.4: Reference graph databases per dataset

3.2.5 Distance approach

When performing graph-based FDI, the Distance approach centres on comparing the total distance difference between two graphs using the attributed values from each graph's respective node signature matrix [15]. The initial matching algorithm presented in [15] was applied in matching object images according to distance similarity between their respective node signatures arranged into graphs. This method was then applied towards a different goal in [4], where this algorithm was implemented to perform fault detection and isolation on an industrial process model. Applying the Distance approach to the graph-matching FDI method has been further utilised in literature, with notable application on detecting fault conditions presented by the gas-to-liquid model by [5] under steady-state conditions, and by [6] on an industrial steam turbine system. While results of this FDI method varied between these studies, taking into account their application towards two different systems, the FDI method still presented notable performance measured in regard to its ability for detecting and isolating conditions, both NOC and fault, as observed from presented datasets [5, 6].

The Distance approach consists of comparing each operational graph series for a respective operation condition to each reference graph in the database of known conditions observed during system operation (NOC and 20 faults). This is to determine which reference graph presents the highest similarity, as per the lowest distance between the graphs compared. This is calculated over the span of the operational graph series S, as the true operating condition represented by the operational graph series is unknown prior to executing the FDI classification by the FDI method. Figure 3.5 demonstrates a high-level overview in comparing the operational graphs of fault 1, with graph series length S, towards each of the respective reference graphs in the database to deliver the distance as the difference between these two graph types over the range S graphs. The resulting distance value for operational graph entry s in S is denoted as $D_{(OP:REF)}(s)$ in respect to both conditions associated with the graph comparison performed. The utilisation of these distance values (thus $D_{(OP:REF)}(S)$ per reference graph considered) when performing graph comparison for the purpose of fault detection and isolation will be discussed during a following sub-section regarding this FDI method.

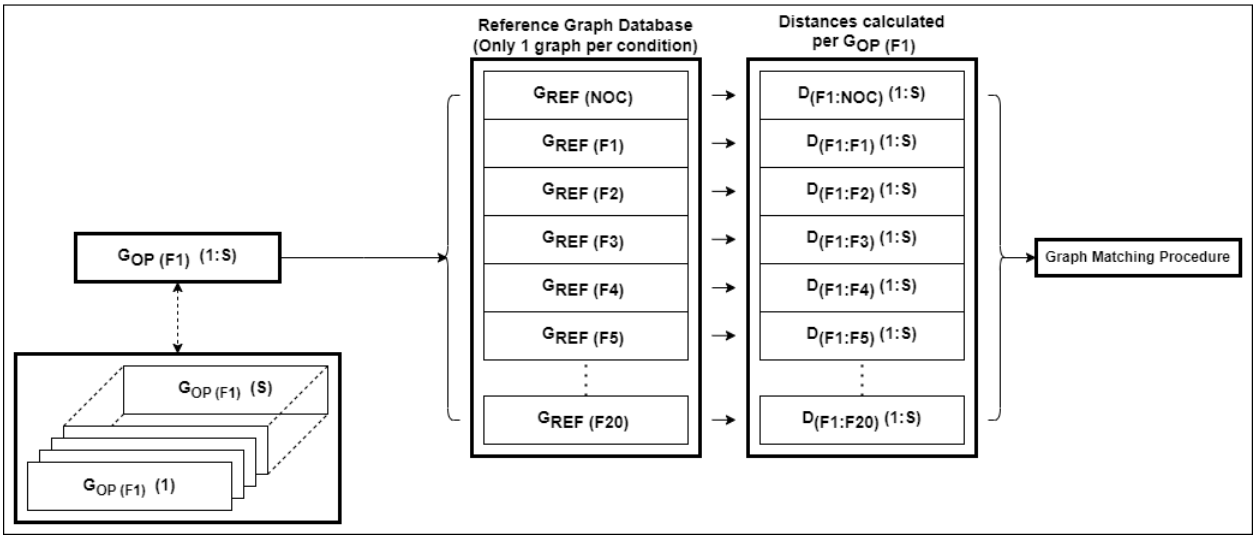

Figure 3.5: Distance approach: Flow of Operational- and Reference graphs

Node distance cost

The distance between two multi-vector data sources, such as operational and reference graphs respectively, can be determined using the Heterogeneous Euclidean Overlap Metric (HEOM), where calculated distance values between attributed nodes would indicate the measure of overlap between the considered sources [27]. The unitless distance delivered by this method provides an indication to the similarity between the considered sources (graphs), where complete similarity would consist of a zero value as distance, and dissimilarity being alternatively represented by an increasing distance value (non-zero). This application of the distance-based approach for graph matching using HEOM follows the methodology presented in [15], where (10) defines the HEOM function with a being an attribute of the number input variables m [27].

$$HEOM(i,j) = \sqrt{\sum_{a=1}^{m} \delta(i_a, j_a)^2} \tag{10}$$

$\delta(i_a, j_a)$ defines the distance variable per attribute type considered during calculation [15]:

$$\delta(i_a, j_a) = \begin{cases} 1 & \text{if } i_a \text{ or } j_a \text{ is missing or unknown} \\ overlap(i_a, j_a) & \text{if } a \text{ is symbolic (qualitative)} \\ rn_{diff_a}(i_a, j_a) & \text{if } a \text{ is numeric (quantitative)} \end{cases} \tag{11}$$

Equation (11) then reduces to only use rn_{diff_a}, since the exergy values delivered by the TEP model are quantitative in nature:

$$\delta(i_a, j_a) = rn_{diff_a}(i_a, j_a)$$

The distance term $rn_{diff_a}(i_a, j_a)$ [27] is further expanded in (12), with $range_a$ responsible for normalising the result to between [0,1].

$$rn_{diff_a}(i_a, j_a) = \frac{|i_a - j_a|}{range_a} \tag{12}$$

By substituting (12) to the associated term in (10), the HEOM implemented during this study can be seen in (13) and (14) respectively.

$$HEOM(i, j) = \sqrt{\sum_{a=1}^{m} \left(\frac{|i_a - j_a|}{range_a} \right)^2} \tag{13}$$

Calculating similarity between two (15×17) graphs requires determining the cost of each node entry within the node signature matrix when calculating the cost value between the operational graph node and the respective reference graph node considered [5]:

$$COST(i, j) = HEOM(i, j) = \sqrt{\sum_{w=1}^{W} \left(\frac{|G_{op}(i, w) - G_{ref}(j, w)|}{range_{ref}(w)} \right)^2} \tag{14}$$

The range in (15) includes w as attribute of W, defined as the width of the considered node signature matrix for the system. The range was calculated using the per column values of the considered reference graph to deliver fixed normalisation values when comparing it to the series of operation graphs [5].

$$range_{ref}(w) = |max(G_{ref}(w)) - min(G_{ref}(w))| \tag{15}$$

Cost matrix

Continuing the process of calculating the distance as cost value between an operational graph node and respective reference graph node, calculating the cost value delivered by (14) must then be repeated for each node entry in the signature matrix to deliver the complete cost matrix with dimension size $n \times n$, as determined by the considered graphs. The cost matrix contains the entire set of calculated distances observed between the two respective graphs [5,6].

$$C(1,1) = \sqrt{\left(\frac{|N_{op}(1,1) - N_{ref}(1,1)|}{range_1}\right)^2 + \left(\frac{|N_{op}(1,2) - N_{ref}(1,2)|}{range_2}\right)^2 + \cdots + \left(\frac{|N_{op}(1,n) - N_{ref}(1,n)|}{range_n}\right)^2}$$

$$C(1,2) = \sqrt{\left(\frac{|N_{op}(1,1) - N_{ref}(2,1)|}{range_1}\right)^2 + \left(\frac{|N_{op}(1,2) - N_{ref}(2,2)|}{range_2}\right)^2 + \cdots + \left(\frac{|N_{op}(1,n) - N_{ref}(2,n)|}{range_n}\right)^2}$$

$$C(2,1) = \sqrt{\left(\frac{|N_{op}(2,1) - N_{ref}(1,1)|}{range_1}\right)^2 + \left(\frac{|N_{op}(2,2) - N_{ref}(1,2)|}{range_2}\right)^2 + \cdots + \left(\frac{|N_{op}(2,n) - N_{ref}(1,n)|}{range_n}\right)^2}$$

Demonstrated in (16) is the full cost matrix delivered when comparing two graphs, with operational graph nodes n and reference graph nodes m defining cost matrix size, and notation $C_{(OP:REF)}$ defining the distance from respective operation graph to current reference graph used.

$$\mathbf{C_{(OP:REF)}} = \begin{bmatrix} C_{1,1} & C_{1,2} & C_{1,3} & C_{1,4} & C_{1,5} & C_{1,6} & C_{1,7} & \cdots & C_{1,m} \\ C_{2,1} & C_{2,2} & C_{2,3} & C_{2,4} & C_{2,5} & C_{2,6} & C_{2,7} & \cdots & C_{2,m} \\ C_{3,1} & C_{3,2} & C_{3,3} & C_{3,4} & C_{3,5} & C_{3,6} & C_{3,7} & \cdots & C_{3,m} \\ C_{4,1} & C_{4,2} & C_{4,3} & C_{4,4} & C_{4,5} & C_{4,6} & C_{4,7} & \cdots & C_{4,m} \\ C_{5,1} & C_{5,2} & C_{5,3} & C_{5,4} & C_{5,5} & C_{5,6} & C_{5,7} & \cdots & C_{5,m} \\ C_{6,1} & C_{6,2} & C_{6,3} & C_{6,4} & C_{6,5} & C_{6,6} & C_{6,7} & \cdots & C_{6,m} \\ C_{7,1} & C_{7,2} & C_{7,3} & C_{7,4} & C_{7,5} & C_{7,6} & C_{7,7} & \cdots & C_{7,m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n,1} & C_{n,2} & C_{n,3} & C_{n,4} & C_{n,5} & C_{n,6} & C_{n,7} & \cdots & C_{n,m} \end{bmatrix} \tag{16}$$

Simplification of the cost matrix instance during [15] includes using the Hungarian method to determine the optimal minimal cost value using all matrix entries. Since the respective graph types implemented during this study present the same system containing identical geometric arrangement of nodes ($n = m$) and their respective interactions, circumstances allow for simplification of the cost matrix using only entries located on the matrix diagonal, as seen in Figure 3.6 where diagonals represent the cost per corresponding node in both graphs considered. These diagonal matrix entries are then arranged as a simplified cost matrix for ease of reference going forward [5].

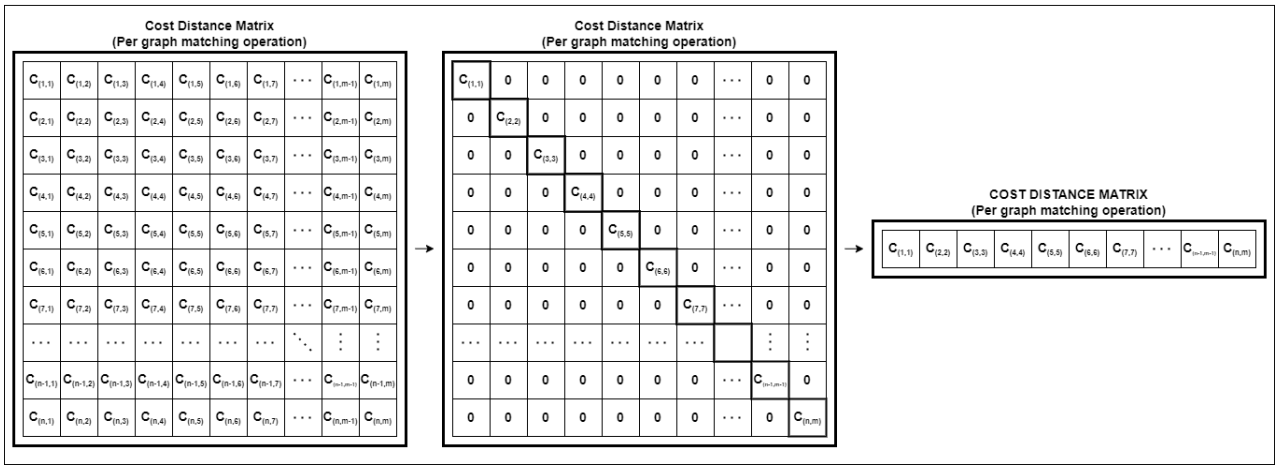

Figure 3.6: Cost matrix simplification

Following the simplification seen during Figure 3.6, the mean value calculated from this simplified cost matrix will represent the single distance value between the two respective graphs compared. The resulting distance, as similarity indicator, from this operation denotes $D_{(OP:REF)}(s)$ as defining the distance from respective operational graph entry s to the reference graph used.

$$D_{(OP:REF)} = \frac{\sum_{i=1}^{n} C_{(i:i)}}{n} \tag{17}$$

In n is the height of the square cost matrix in Figure 3.6 and i indicates the diagonal entry position. These calculated distance values would then be used to perform detection and isolation of operation conditions observed, as demonstrated in the subsection that follows.

Distance utilisation for FDI

The graph matching procedure presented thus far only forms part of the FDI process where further utilisation of the calculated graph distances would determine the operation condition detected and isolated if unique. The methodology implemented by [6] for this respective approach considers a series of operational graphs compared to each reference graph in the database individually. Comparing the series S of operational graphs with a reference graph will deliver a series of distance values $D_{OP:REF}(S)$ for each reference graph compared: in this case 21 sets of distance value series for this study. These distances will then be compared in a row-wise manner per position s for the minimum value, with Figure 3.7 illustrating an example to this layout during comparison. The column position, indicating the respective reference graph, and demonstrating the smallest distance value per row would thus indicate the condition to which the operational graph demonstrates the highest similarity.

NOC	FAULT 1	FAULT 2	FAULT 3	FAULT 4	----------	FAULT 20
$D_{(F1:NOC)}(1)$	$D_{(F1:F1)}(1)$	$D_{(F1:F2)}(1)$	$D_{(F1:F3)}(1)$	$D_{(F1:F4)}(1)$	----------	$D_{(F1:F20)}(1)$
$D_{(F1:NOC)}(2)$	$D_{(F1:F1)}(2)$	$D_{(F1:F2)}(2)$	$D_{(F1:F3)}(2)$	$D_{(F1:F4)}(2)$	----------	$D_{(F1:F20)}(2)$
$D_{(F1:NOC)}(3)$	$D_{(F1:F1)}(3)$	$D_{(F1:F2)}(3)$	$D_{(F1:F3)}(3)$	$D_{(F1:F4)}(3)$	----------	$D_{(F1:F20)}(3)$
$D_{(F1:NOC)}(4)$	$D_{(F1:F1)}(4)$	$D_{(F1:F2)}(4)$	$D_{(F1:F3)}(4)$	$D_{(F1:F4)}(4)$	----------	$D_{(F1:F20)}(4)$
$D_{(F1:NOC)}(5)$	$D_{(F1:F1)}(5)$	$D_{(F1:F2)}(5)$	$D_{(F1:F3)}(5)$	$D_{(F1:F4)}(5)$	----------	$D_{(F1:F20)}(5)$
⋮	⋮	⋮	⋮	⋮	⋱	⋮
$D_{(F1:NOC)}(S)$	$D_{(F1:F1)}(S)$	$D_{(F1:F2)}(S)$	$D_{(F1:F3)}(S)$	$D_{(F1:F4)}(S)$	----------	$D_{(F1:F20)}(S)$

Figure 3.7: Example distance entries per G_{F1}

These row-wise minimum values are then recorded as to their column position, indicating likeliest reference graph match, and summed per column over the length of operational graphs in the series S to deliver the total graphs matched to each reference graph. The sum total per column could then be expressed as a percentage by dividing by S to illustrate the share of operational graphs found demonstrating the similarity of each respective reference graph; i.e., demonstrating similarity to the respective operation conditions associated with this system. The reference graph demonstrating the largest share of similar operational graphs would then become identifiable as the condition observed by the respective FDI method.

Fault classification is performed as discussed in Chapter 2 where outcomes are based on boolean logic when comparing the resulting column values in Figure 3.7. Figure 3.8 demonstrates the fault detection classification for the normal operating condition, accompanied by Figure 3.9, which additionally demonstrates the respective classification for the fault conditions when using the same array of similarity allocations per reference graph. The fault classification presented in Figure 3.9 is valid for all fault conditions, where individual isolation would further be determined by the position of the largest similarity allocation present in this array.

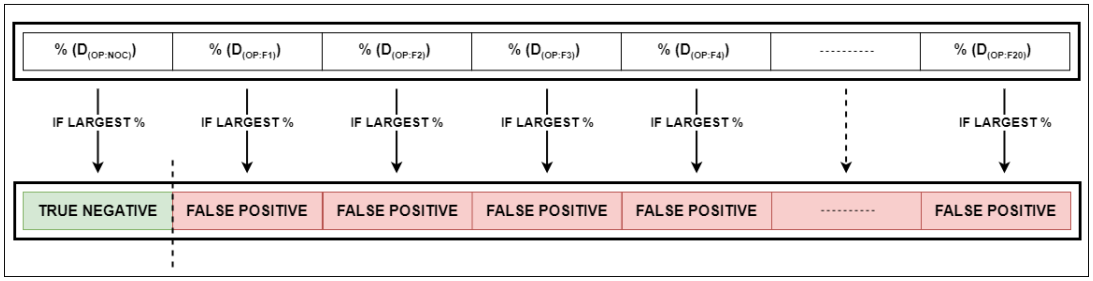

Figure 3.8: Confusion matrix classification of NOC operational graph: Distance approach

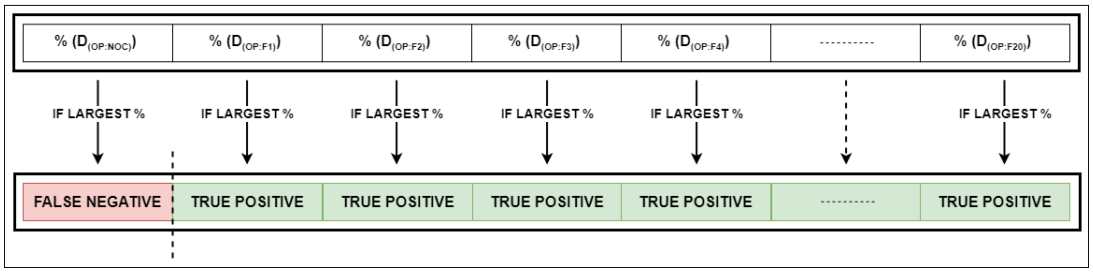

Figure 3.9: Confusion matrix classification of fault operational graphs: Distance approach

Following the condition classifications presented in Figures 3.8 and 3.9 for the respective operation graphs per condition type, the array of graph similarity values for each respective operational graph series S would then be arranged in a table format, starting with the NOC and followed by each successive respective fault condition. The operational graph order then matches the reference graph database order, with diagonal entries representing the corresponding operating condition.

Figure 3.10 demonstrates the subsequent graph comparison results delivered when comparing all 21 operation conditions with the reference graph database (also 21 conditions), where each row demonstrates the share of similarity established by the operational graph series towards the respective reference graph entries. A correct isolation of an operating condition then requires the detected condition classification to be located on the respective diagonal entry, in a corresponding reference graph database of the results table. An operating condition is unisolable, should more than one condition be classified as a true condition per row.

	$G_{REF(NOC)}$	$G_{REF(F1)}$	$G_{REF(F2)}$	$G_{REF(F3)}$	$G_{REF(F4)}$	.....	$G_{REF(F20)}$
$G_{OP(NOC)}$	TRUE NEGATIVE	FALSE POSITIVE	FALSE POSITIVE	FALSE POSITIVE	FALSE POSITIVE	.....	FALSE POSITIVE
$G_{OP(F1)}$	FALSE NEGATIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	.....	TRUE POSITIVE
$G_{OP(F2)}$	FALSE NEGATIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	.....	TRUE POSITIVE
$G_{OP(F3)}$	FALSE NEGATIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	.....	TRUE POSITIVE
$G_{OP(F4)}$	FALSE NEGATIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	.....	TRUE POSITIVE
$G_{OP(F5)}$	FALSE NEGATIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	.....	TRUE POSITIVE
$G_{OP(F6)}$	FALSE NEGATIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	.....	TRUE POSITIVE
$G_{OP(F7)}$	FALSE NEGATIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	.....	TRUE POSITIVE
⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮
$G_{OP(F20)}$	FALSE NEGATIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE	TRUE POSITIVE

Figure 3.10: Confusion matrix classification of operation graphs per dataset

3.2.6 Eigenvalue decomposition - Qualitative approach

The Eigendecomposition approach performs graph matching, by comparing deviation in scalar eigenvalues of the calculated self-cost matrix of a reference graph and the cost matrix determined between each operational graph, with the respective reference graph matching calculated. Evaluation of these deviations for fault detection purposes could either be done qualitatively or quantitatively, where this section will focus on the Qualitative approach. The Eigendecomposition-Quantitative approach is discussed in the section to follow.

Self-cost matrix

This approach demonstrates methodology common with the Distance approach where calculation of the distance cost matrix between two considered graphs in (14) delivers only the initial component as part of the comparison process performed during this approach. Eigendecomposition of an $n \times n$ matrix A with non-zero eigenvector $\mathbf{x}$ such that (18) holds for the scalar λ. This λ is the eigenvalue of corresponding eigenvector $\mathbf{x}$. These components to the transformation in (18), respectively demonstrate the associated transform magnitude and orientation of the vectors with negative vector values indicating opposing direction [32, 33].

$$A\mathbf{x} = \lambda\mathbf{x} \tag{18}$$

To perform an eigenvalue comparison during this approach, each respective reference graph is first compared with itself to deliver a self-imposed cost matrix, where eigendecomposition of this matrix delivers the respective eigenvalues and eigenvectors as a fixed reference point from which graph similarity could be observed [4, 5]. Sole use of the eigenvalues was considered for this study based on the previous application observed in [5] and [6], rather than to the comparison performed by [33] where eigenvectors where included as part of the fault signature by which matching was performed. The resulting self-cost matrix delivered when comparing a reference graph with itself is presented in Figure 3.11 (left). Eigendecomposition is then performed using this matrix to deliver its respective eigenvalues and eigenvectors. Only the eigenvalues are used during this FDI method where values are arranged in a row matrix λ_{REF} in Figure 3.11 (right).

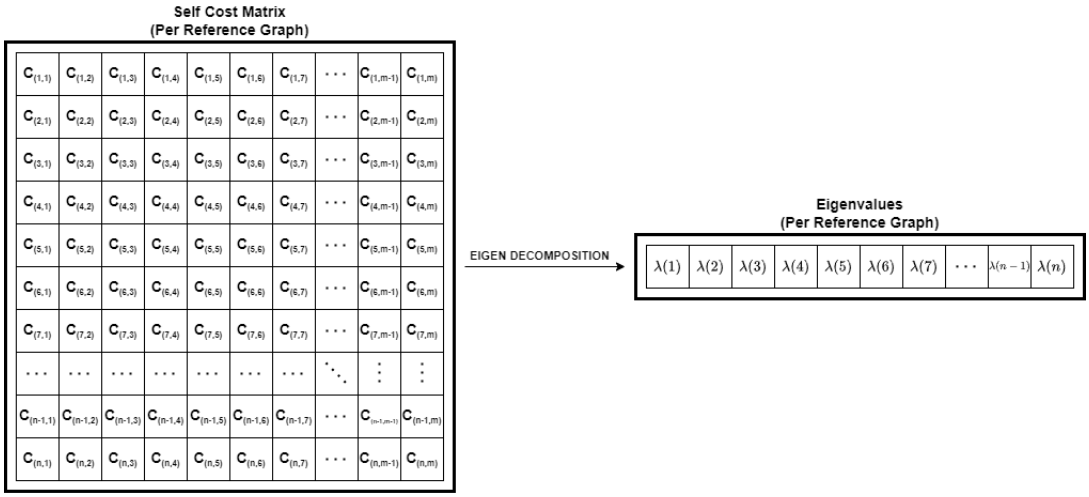

Figure 3.11: Self-cost matrix during a qualitative approach

The graph matching procedure progresses using the same initial methodology as the Distance approach, by implementing the HEOM method to first compare the reference graph with itself to deliver a self-cost matrix. Eigendecomposition of this matrix then delivers the set of eigenvalues used as a fixed reference point from which graph similarity would be determined below. Next, the distance between the operational graph and the respective reference graph is calculated as done during the Distance approach. The resulting cost matrix is further decomposed to deliver n eigenvalues arranged as row matrix with each column entry denoted $\lambda_{OP:REF}(n)$ with OP operational graph and REF reference graph used during comparison. This procedure is repeated for each operational graph in the considered series to deliver a series of eigenvalue sets matching series entries as demonstrated in Figure 3.12 for operational graph entries $G_{OP}(1:S)$

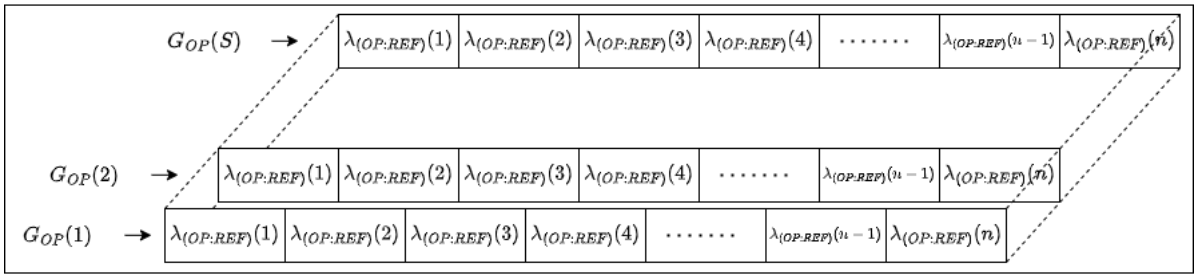

Figure 3.12: Graph comparison eigenvalues from operational graph series

Qualitative eigenvalue utilisation for FDI

The methodology of performing graph comparison for the Qualitative approach consists of evaluating the delivered eigenvalue cost matrices, λ_{REF} and $\lambda_{OP:REF}$, to determine whether the calculated similarity per eigenvalue entry is within a defined threshold value. Previous implementation of this method includes the utilisation of the three standard deviation rule (3σ) as threshold, from which outliers would be determined. Implementation of the threshold for qualitative assignment follows [5] in using 3 standard deviation values as threshold, from which a classification in the qualitative assignment was performed. Equation (19) presents the condition from which the assignment was performed when evaluating for possible outliers from the calculated eigenvalue cost difference per graph comparison case. Figure 3.13 is an illustration of how (19) would be considered during each entry of the eigenvalue matrix used during graph comparison [5].

$$T_{op:ref}(n) = \begin{cases} 0 & if|\lambda_{ref}(n) - \lambda_{(op:ref)}(n)| < 3\sigma \\ 1 & otherwise \end{cases} \tag{19}$$

$\lvert\lambda_{(REF)}(1) - \lambda_{(OP:REF)}(1)\rvert$	$\lvert\lambda_{(REF)}(2) - \lambda_{(OP:REF)}(2)\rvert$	$\lvert\lambda_{(REF)}(3) - \lambda_{(OP:REF)}(3)\rvert$	.....	$\lvert\lambda_{(REF)}(n) - \lambda_{(OP:REF)}(n)\rvert$

Figure 3.13: Qualitative assignment of eigenvalue difference

The standard deviation (SD) values per respective position for the eigenvalue cost matrix were calculated using the equation demonstrated during [5], whereas the standard deviation was calculated using the self cost-matrix entries for each reference graph. This delivered an individual threshold for each of the reference graphs considered. Details of how this procedure was implemented for this study is further demonstrated in Appendix A of this document. The resulting set of qualitative values assigned are then summed as the total similarity demonstrated by the considered operational graph, as entry of a series, to a reference graph in respect of the condition represented. A lower difference value demonstrated by this sum of qualitative assigned positions would indicate a higher degree of similarity, while the maximum value, represented by the number of graph nodes in the system considered, indicates the least similarity per comparison case.

Interpretation of Qualitative approach results

After performing this qualitative assessment which resulted in a single value indicator of similarity between an operational graph and a reference graph, the results need to be interpreted when each entry of an operational graph series is compared with such a series associated with each operating condition. This is then compared sequentially through each reference graph database entry. The sum of qualitative assignments when comparing $G_{OP}(n)$ and $G_{REF}(m)$ remain as a series of values and is stored in the position (n, m) of a matrix, noting the encompassing space when comparing 21 operational graph series with 21 reference graphs in the constructed database, as illustrated in Figure 3.14.

Qualitative values for operational graphs, demonstrated as the rows in Figure 3.14, used in reference graphs in the database are compared simultaneously when flitting through entries in the respective series N of qualitative results. In considering an operational graph row from Figure 3.14, the qualitative values contained in columns for position n, as entry of series N, will be compared side by side to determine the lowest value per n. The column position demonstrating this lowest value in position n of the series is allocated a point which is indicative of demonstrating the highest similarity (likeliest condition). This is repeated for each entry in series N for the respective operation graph, where the total points per column is then demonstrated as a percentage of entries in N. This indicates the share at which each reference graph condition was matched as the likeliest match to the condition demonstrated by the operational graph observed. Figure 3.15 demonstrates an example of the above process, as considered for the fault 8 operational graph series across reference graphs

in the constructed database.

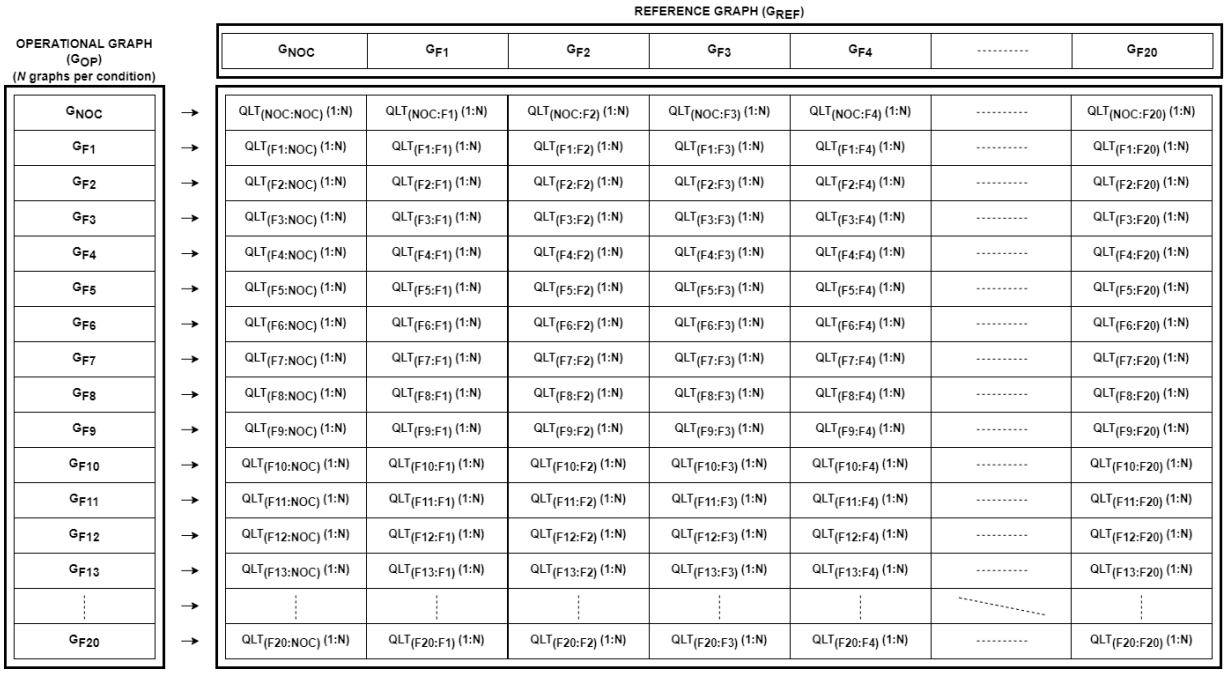

Figure 3.14: Qualitative comparison tracking between operational and reference graphs

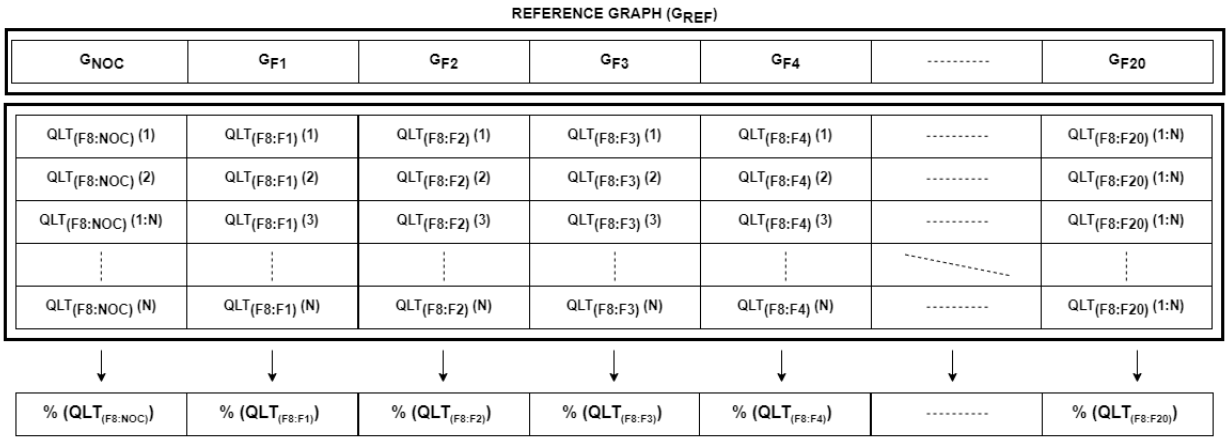

Figure 3.15: Ratio of qualitative assignments of an operational graph series: Fault 8

Lastly, the table of results obtained from the above process would then be evaluated in a similar manner to the confusion matrix demonstrated for the Distance approach. The normal operating condition is evaluated similarly to the Distance approach, as demonstrated in Figure 3.16, which is followed by Figure 3.17 illustrating the case for evaluating all fault conditions considered. The summary illustration for the fault detection confusion matrix, as seen in Figure 3.10 above, remains as applicable as with this approach, where the same case is made to isolating fault conditions. The row entry demonstrating the largest ratio allocated after the comparison process would then be classified as the isolated condition by the FDI method used.

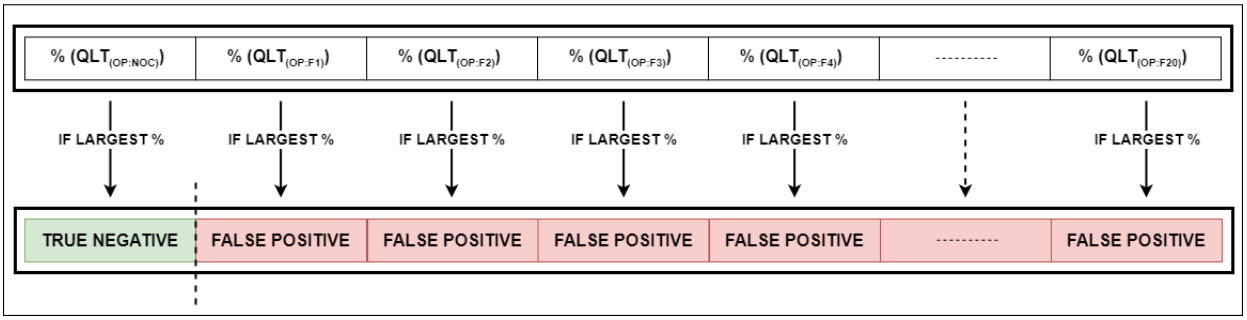

Figure 3.16: Confusion matrix classification of NOC operational graph: Quantitative approach

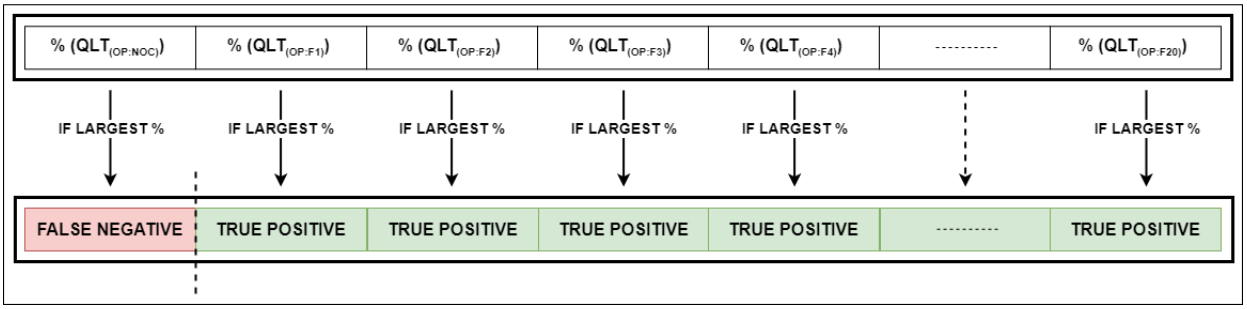

Figure 3.17: Confusion matrix classification of fault operational graphs: Quantitative approach

3.2.7 Eigenvalue decomposition - Quantitative approach

Following the methodology presented above, the Quantitative approach follows the same methodology as that of the Eigenvalue-Qualitative approach, from the initial process to calculating an eigenvalue-based graph difference. A summary of these steps shared with the above-mentioned Qualitative approach are summarised below:

- Calculate the self-cost matrix of the respective reference graph.

- Perform eigendecomposition of this resulting self-cost matrix, obtaining the eigenvalues from this transform.

- Determine the distance between operational and reference graphs for each operational graph in the series.

- Perform eigendecomposition of this operational & reference graph cost matrix per operational graph in the series.

- Calculate the difference per eigenvalue position between eigenvalues resulting from the self-cost matrix obtained above and the eigenvalues from the operational & reference graph cost matrix. This is repeated for each operational graph in the considered series.

This last step then corresponds with that demonstrated in Figure 3.13 seen above for the Qualitative approach. The deviation between the two eigendecomposition-based approaches now presents itself where, instead of storing a single similarity value (through qualitative assignments) as during the Qualitative approach, the Quantitative approach stores the sum of all difference values per the respective operational graph entries in the considered series.

Interpretation of Quantitative approach results

Following the above steps, the scope of operations performed when comparing each operational graph series accordingly with each reference graph in the constructed reference graph database will follow the same structure as previously seen in Figure 3.14. Each matrix position consists of N entries as per operational graph series length, one per operational-reference graph comparison performed, with the quantity of difference values matching the number of system nodes. Evaluation of these matrix entries towards fault detection, and further fault isolation, is done by comparing entries in the operational graph series concurrently across the reference graph columns. Figure 3.18 demonstrates how one operational graph in position s is illustrated to the respective quantitative difference values, with total system nodes n, where the row-wise minimum (as per eigenvalue position) is evaluated. The position of each row-wise

minimum is allocated a point, with the sum representing the total similarity demonstrated by the considered operational graph up to the reference graph per column. Lastly, the sum of classifications per reference graph column is considered per the operational graph series S length, as demonstrated during both the Distance and Qualitative approaches above.

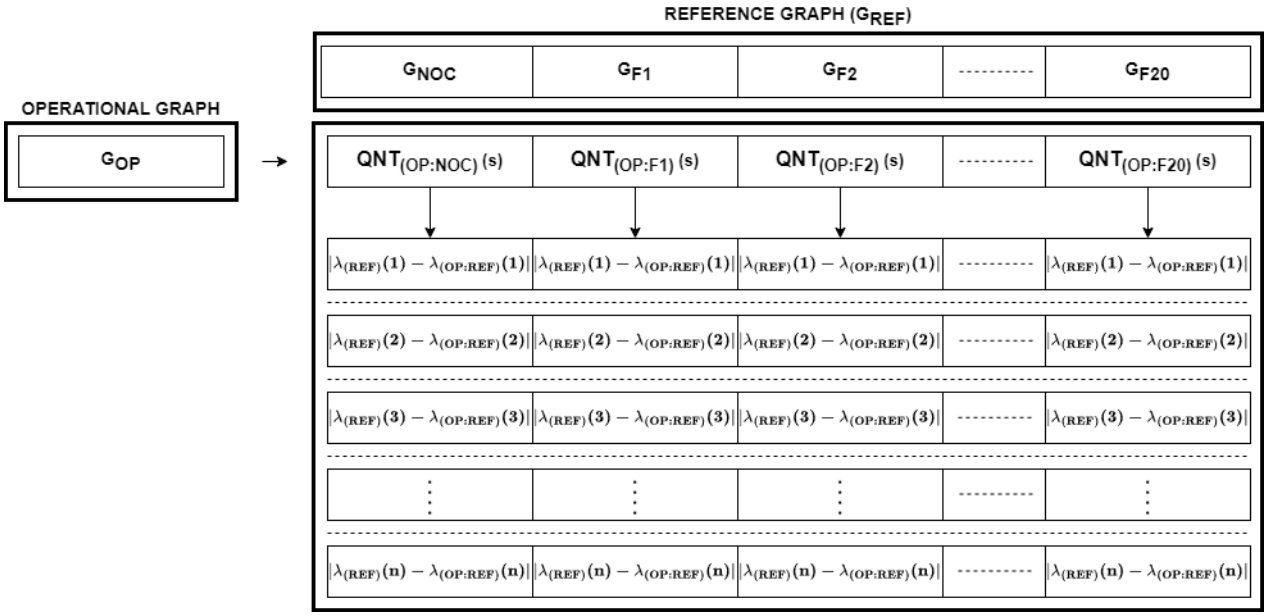

Figure 3.18: Eigenvalue difference matrix per reference graph comparison

Condition classification would be performed using the established confusion matrix, as previously presented during the above approaches in Figures 3.8, 3.9 and 3.10 respectively, where detection results would be contained in the square matrix with the length corresponding to the number of system nodes. Isolation of conditions observed would correspond per diagonal entry to the matrix which is illustrated for valid classification, while any off-diagonal allocation would result in an unisolable classification.

3.3 EGBV method functional parameters

3.3.1 Functional parameter locations

Literature surveyed on the implementation of the above EGBV methods in performing FDI allowed the researcher to identify key locations where functional parameter selection could occur. Locations were identified to be commonly shared between the three identified EGBV methods, with selectable options varying between the respective EGBV methods during each functional parameter location.

Distance function

The distance function as discussed during the above EGBV methods is the means by which discernible difference is calculated when comparing attributes, as would be used when comparing points between graphs, to deliver a single value as operation result [27, 28].

Normalisation

Normalisation serves the function of scaling the input attributes proportionately to reduce the effect of outliers observed during operation, while allowing values to be scaled into a manageable/fixed range [27, 28].

Cost-matrix utilisation

Cost-matrix utilisation, as identified from the literature surveyed, reflects how the diagonal cost matrix entries, when performing the HEOM, are included during calculation of the graph comparison as a final similarity indicator between considered graphs.

A summary of the functional parameter locations, illustrated in Figure 3.19 below, demonstrates how the respective functional parameter options identified in each location will be implemented when exploring the various functional parameter permutations later in this study.

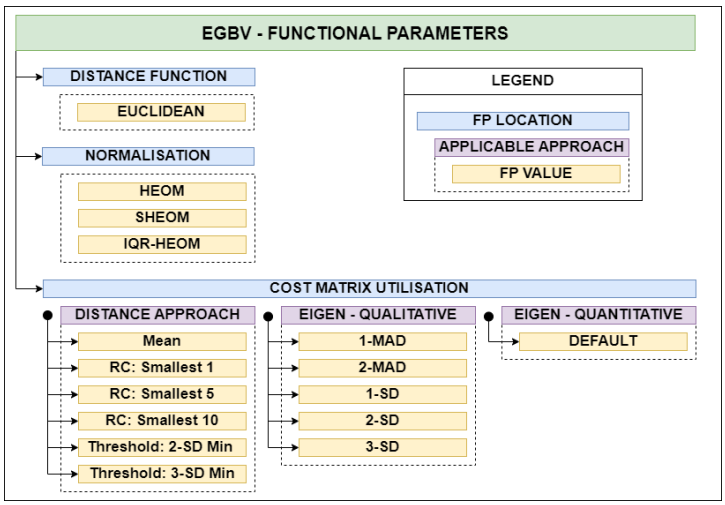

Figure 3.19: EGBV Functional Parameters

Methodology constraint - Distance function

The distance function was identified as a functional parameter location for the role it serves during execution of the EGBV methods during FDI application. Literature by [27] presents the availability of multiple alternate functions that could serve as functional parameter options in this functional parameter location. A crucial difference between the default Euclidean distance function and other distance function exists per the following:

- The Euclidean distance function allows for calculating either single or multi-axis distances. The single axis distance calculation is a simpler form, where distance is calculated as a direct line between two considered points or attributes. This serves to be the most applicable option in the context of this study, where the node-signature matrix is populated with only single values per matrix cell [27, 28].
- Alternate distance functions, such as the Minkowski or Manhattan distance function, can solely function through the use of two or more axes in calculating distance between two points or attributes [27].

The cell entries present in the node-signature matrix used during FDI only contain a single value representing solely one dimension by which distance could be evaluated in this study. In the case of using the Manhattan distance, the function simplifies to a single axis calculation as with the Euclidean distance and would not suitable for this study. Due to the limitation imposed by the current node-signature matrix structure only implementing single values per cell, a constraint is placed on this functional parameter location to only accommodate the Euclidean distance function as a suitable option. In the case where cost matrix entries contain 2 or more values per cell then alternate distance functions would be applicable. The exploration of alternate options in arranging the node-signature matrix to include multiple values per cell

entry is beyond the scope of this study. This leads to the distance function being identified as a functional parameter location, while simultaneously acknowledging that this location will not be investigated due to the constraint placed on the current node-signature matrix only containing single value cell entries. The Euclidean distance function will be kept as default functional parameter value during all three EGBV methods implemented during this study.

3.3.2 Functional parameters - Distance approach

The functional parameter locations identified in the preceding sub-section are now discussed per context of the Distance approach, with the applicable functional parameter options per location as seen in Figure 3.20 below.

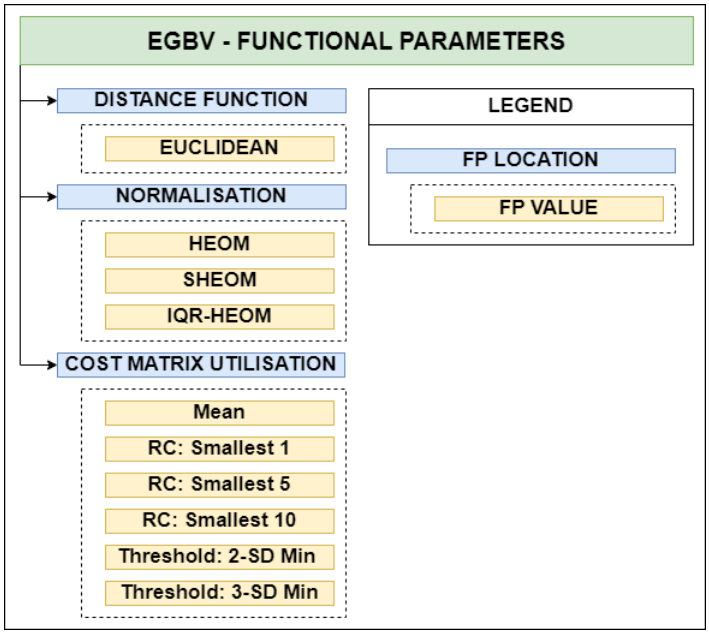

Figure 3.20: Functional parameters - Distance approach

Functional parameter options - Normalisation

Advancements have been presented on the HEOM in [28], where an alternate method to normalisation during the HEOM was implemented using standard deviation of the attributes considered [27]. An additional alternative to this normalisation was demonstrated by [29], where the interquartile range was used instead to reduce the impact of outlier input values considered. Variations between these 3 instances are presented below:

1. HEOM (Default)

The default normalisation value implemented in the HEOM uses the range of considered attribute values as presented in (15) earlier in this chapter.

$$range_{ref}(w) = |max(G_{ref}(w)) - min(G_{ref}(w))|$$

2. SHEOM

The Scaled Heterogeneous Euclidean Overlap Metric (SHEOM) [28] differs by instead using the standard deviation in normalising attribute values when calculating the distance, as displayed in (15) with the Manhattan distance being implemented by [28],

$$SHEOM(x_i, x_j) = \sum_{r=1}^{R} \frac{|x_{i,r} - x_{j,r}|}{s_r} \tag{20}$$

Where adaptation in the context of this study was made using the Euclidean distance, with the column-wise standard deviation of the operational graph in (21), and (22) respectively delivering the SHEOM with notation substituted [28]:

$$s_{ref}(w) = \sqrt{\frac{\sum_{n=1}^{N} \left(G_{ref}(n, w) - \overline{G_{ref}(w)}\right)^2}{N}} \tag{21}$$

$$SHEOM(i, j) = \sqrt{\sum_{w=1}^{W} \left(\frac{|G_{ref}(i, w) - G_{op}(j, w)|}{s_{ref}(w)}\right)^2} \tag{22}$$

3. IQR-HEOM

The IQR-HEOM by [29] is a similar statistical approach to normalisation as the SHEOM, by using collective attribute values considered during the operation. Equation (23) presents how the statistical interquartile range positions would serve as the normalisation values in (24), with the full substitution to respective graph types [29].

$$iqr_{ref}(w) = Q_3(w) - Q_1(w) \tag{23}$$

$$IQR - HEOM(i,j) = \sqrt{\sum_{w=1}^{W} \left(\frac{|G_{ref}(i,w) - G_{op}(j,w)|}{iqr_{ref}(w)} \right)^2} \tag{24}$$

Functional parameter options - Cost-matrix utilisation

The utilisation of the cost matrix was identified as functional parameter location, where the opportunity exists to selectively implement entries in providing an alternate solution when interpreting cost values from the considered range of graph nodes. This creates an opportunity to observe any sensitivity demonstrated towards outliers, while also exploring whether a dampening effect could be achieved to any outliers observed in the simplified cost matrix.

1. Mean (default)

The default method in reducing the cost matrix is done by calculating the mean value by using all entries of the simplified cost matrix, illustrated earlier in Figure 3.6, to establish a single value indicator for graph similarity [27].

2. Reduced capacity (RC): Smallest 1

This functional parameter option serves to demonstrate the outcome by basing the graph similarity outcome solely on a reduced capacity of the simplified cost matrix. This case only considers one node in the respective simplified cost matrix, demonstrating the highest similarity resulting from the graph comparison performed. This serves in establishing the possibility of identifying an operating condition by the single closest node match.

3. Reduced capacity (RC): Smallest 5

This option expands on the case stated during the above entry, where the reduced capacity of node values used from the simplified confusion matrix is expanded to include an increased portion of the total entries. A more sizeable quantity of nodes was selected as reduced capacity, in order to further explore any changes to FDI capability and/or dampening of outlier values that this may have. As the TEP system used during this study consists of 15 nodes, a reduced capacity value of 5 (from 15) was selected for this interval, representing 33% total node capacity of the simplified cost-matrix. This would additionally reinforce how much FDI performance, in comparison to the default method, remains consistent when using such a reduced capacity of the simplified cost-matrix.

4. Reduced capacity (RC): Smallest 10

This option, as an extension of the previous, further explores the above argument using the minor reduction in capacity of the simplified cost-matrix entries, using 10 (from 15) node entries towards representing 66% of the nominal quantity.

5. Threshold: 2 Standard deviation from minimum (2-SD Min)

This functional parameter option follows a different route, with only utilising matrix entries within a determined threshold in an attempt to eliminate outlier values that could otherwise influence the calculated final similarity indicator. The standard deviation (σ) is calculated from the simplified cost matrix, followed by identifying the minimum value entry. Each simplified cost matrix entry is evaluated to determine whether it's value falls within 2 standard deviations from the minimum matrix entry; all matrix entries within this threshold are summed towards a total value which is to be used as the final similarity indicator. This option serves to remove outliers observed while reducing the likelihood of a masking effect caused by using a mean-value-based similarity indicator.

6. Threshold: 3 Standard deviation from minimum (3-SD Min)

This functional parameter option further extends the observable threshold as to implement a more stringent criteria towards the classification of outliers in confusion matrix entries.

3.3.3 Functional parameters - Qualitative approach

The functional parameter locations identified for the Eigendecomposition - Qualitative approach are discussed below, with applicable functional parameter options for each location defined in Figure 3.21.

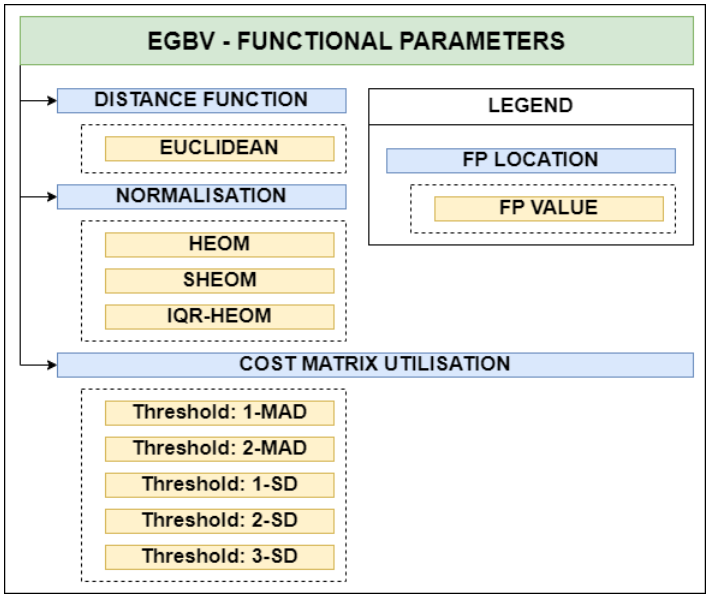

Figure 3.21: Functional parameters - Qualitative approach

Functional parameter options - Normalisation

Implementation of the HEOM, SHEOM and IQR-HEOM functional parameter options will follow the same methodology as presented during the discussion of the Distance approach in the previous section.

Functional parameter options - Cost-Matrix Utilisation

The cost-matrix utilisation considered for this approach differs from the previously demonstrated approaches, by rather considering how a threshold value is used in selecting entries applicable to indicating graph similarity. The use of this threshold selection seen during [5] and [6] are based on the statistical probability of outlier occurrence increasing with each addition of the standard deviation from the mean. The following functional parameter options were implemented as variation towards exploring the impact demonstrated when varying the selection threshold, as well as the outcome this presents on method results:

1. 3σ (3-SD)
2. 2σ (2-SD)
3. 1σ (1-SD)

An alternate statistical measure in classifying outliers is through the use of the Mean Absolute Deviation (MAD), which aims to be more robust to outlier values during calculation. The MAD is composed of the absolute sum difference between each value in the considered series and the series mean. The following increments were implemented as options:

4. 2-MAD
5. 1-MAD

3.3.4 Functional parameters - Quantitative approach

The functional parameter locations, as identified for the Eigendecomposition–Quantitative approach, are discussed below, with applicable functional parameter options for each location defined in Figure 3.22.

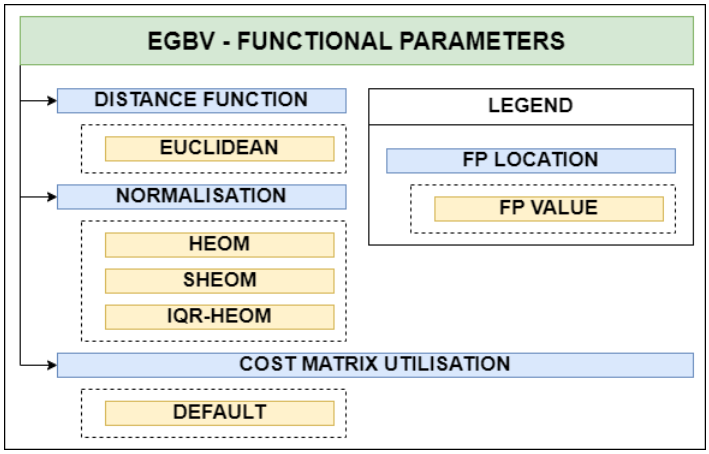

Figure 3.22: Functional parameters - Quantitative approach

Due to the Quantitative approach's utilisation of the eigenvalue cost-difference values as individual measures, the opportunity for alternate functional parameter options was restricted. Deviation from this approach's methodology would result in the approach transitioning into either one of the previously demonstrated approaches. If the Quantitative approach was to evaluate the eigenvalue cost difference matrix as a collection of values, then the cost-matrix utilisation options demonstrated for the Distance approach would become applicable. Such a result in this respective approach would become a hypothetical "Eigenvalue-Distance" approach, which would not serve to add value to the scope of EGBV methods considered during this study.

Functional parameter options - Normalisation

The default methodology of the Quantitative approach during initial cost value calculation provides the opportunity to implement the same normalisation options as previously presented during both Distance and Qualitative approaches:

- HEOM (Default)
- SHEOM
- IQR-HEOM

3.4 Functional parameter permutations

The functional parameter options described above are illustrated in Figure 3.23 where the quantity of permutations is listed for each of the respective FDI approaches considered. A limitation is imposed on the distance function as functional parameter option for this study due to the attributed graph for the TEP model only containing single values per entry. This presents the distance in a single dimension whereas the alternate distance functions identified all rely on distance being measured in 2- or 3 dimensions in order to calculate distance.

	DISTANCE APPROACH	EIGENDECOMPOSITION - QUALITATIVE	EIGENDECOMPOSITION - QUANTITATIVE
DISTANCE FUNCTION	1	1	1
NORMALISATION	3	3	3
COST MATRIX UTILISATION	6	5	1
TOTAL	18	15	3

Figure 3.23: Functional parameter Summary

The Distance approach presents 18 permutations by using the 3 functional parameter locations, as seen in Figure 3.24 which demonstrates the options solely applicable to this approach as illustration of how these permutations expand from functional parameter locations. A similar breakdown of permutations existing for the Qualitative approach is provided in Figure 3.25 with 15 permutations, and further accompanied by Figure 3.26 which illustrates the 3 permutations that will be utilised for the Quantitative approach.

HEOM MEAN	SHEOM MEAN	IQR-HEOM MEAN
HEOM RC: SMALLEST 1	SHEOM RC: SMALLEST 1	IQR-HEOM RC: SMALLEST 1
HEOM RC: SMALLEST 5	SHEOM RC: SMALLEST 5	IQR-HEOM RC: SMALLEST 5
HEOM RC: SMALLEST 10	SHEOM RC: SMALLEST 10	IQR-HEOM RC: SMALLEST 10
HEOM Threshold: 2-SD MIN	SHEOM Threshold: 2-SD MIN	IQR-HEOM Threshold: 2-SD MIN
HEOM Threshold: 3-SD MIN	SHEOM Threshold: 3-SD MIN	IQR-HEOM Threshold: 3-SD MIN

Figure 3.24: Functional parameter permutations - Distance approach

HEOM 3-SD	SHEOM 3-SD	IQR-HEOM 3-SD
HEOM 2-SD	SHEOM 2-SD	IQR-HEOM 2-SD
HEOM 1-SD	SHEOM 1-SD	IQR-HEOM 1-SD
HEOM 2-MAD	SHEOM 2-MAD	IQR-HEOM 2-MAD
HEOM 1-MAD	SHEOM 1-MAD	IQR-HEOM 1-MAD

Figure 3.25: Functional parameter permutations - Qualitative approach

HEOM	SHEOM	IQR-HEOM

Figure 3.26: Functional parameter permutations - Quantitative approach

3.5 Concept validation

Validating the designed concept towards addressing the identified problem serves as an early intervention point towards establishing whether the concept will be sufficient for this study or if the proposed concept requires re-evaluation. This includes the validation that any assumptions and theories underlying the concept are correct and reasonable for the intended purpose [26].

From the problem statement identified during this study, the question proposed in validating the performance of a sensitivity analysis on the EGBV methods is as follows:

Do EGBV methods present sensitivity towards functional parameters?

Inspecting the proposed concept of sensitivity analysis of EGBV methods and the use of functional parameters during this sensitivity analysis presents the following motivation:

Investigating functional parameters towards both location and options available for use during each approach would present an opportunity to further understand the nature of influence attributed to this concept in an effort to answering the identified problem statement. Evaluating how sensitivity is demonstrated would assist towards considering the state of development of these FDI approaches. Furthermore, any notable sensitivity demonstrated would provide grounds for further method refinement, while a lack of sensitivity of the functional parameter changes can give an indication to which default methodology was used. This can in turn demonstrate which advanced development requires minimal refinement (if applicable).

3.6 Conclusion

This chapter presented a discussion on executing each of the 3 FDI approaches, coupled with how each respectively utilised the graph similarity which was calculated using both operational and reference graph types. A further introduction to the concept of functional parameters is presented, followed by further details on where these could occur, as well as the options available in each location as applicable to each of the considered FDI approaches. Lastly, a summary of the permutations for each of the FDI approaches is demonstrated in using combinations of each identified functional parameter. The chapter to follow will then discuss the experimental design and utilises the TEP simulation model in obtaining the required datasets that would be used in Chapter 5 of this study. The MATLAB code developed for the respective EGBV method permutations can be found in Appendix D where a hyperlink to the files is provided.

Chapter 4

Data Compilation

4.1 Introduction

Chapter 3 presented the identification of the relevant EGBV functional parameters in addressing the first research objective for this study. This chapter will serve to answer the next research objective of "Data compilation" presented in Chapter 1.

This chapter will initially inspect data delivered by executing the Tennessee Eastman process simulation model, followed by the identification of experimental variables, as either independent or dependent, which are associated with the model data. Control and execution of the simulation model is then presented, accompanied by a discussion on the structure implemented when executing the model as the datasets are acquired for FDI, by using the EGBV methods presented later in this study. Lastly, this utilisation of the transformed data towards performing FDI using the respective EGBV methods is presented to clarify how the data are used in the next chapter.

4.2 Tennessee Eastman process data

The TEP simulation model implemented in this study was developed by [12] using MATLAB software, with Simulink as extension of this software package, allowing for the modernisation of the original FORTRAN model by [7]. Implementation of the TEP model within the MATLAB and Simulink environment provides users the opportunity for improved interaction with model properties. This includes providing open access to the hierarchy of process components, variables, and underlying sub-systems of the TEP model.

Process data from the TEP simulation model is captured as time series within the MATLAB workspace for the duration of model operation. The TEP process data of individually

measured variables in their initial form present information on the current state observed for the various sub-processes, associated with the five main sub-components or material streams throughout the TEP model. These individual measurements do not immediately contribute information on the energy and exergy-flow required in performing FDI using the respective EGBV methods for this study.

This section will inspect the execution of the TEP simulation model to identify experimental variables that impact simulation model operating conditions, followed by a further discussion on the structure by which the simulation model would be executed in dataset acquisition.

4.2.1 Variable identification

While the execution of the TEP simulation model does not serve as focus point of this study, it does serve a supportive role by being the source of data to be used for FDI and sensitivity analysis of the functional parameters. The nature of datasets delivered by the simulation model would thus also need to be inspected to identify the experimental variables in order to gain perspective on how dataset quality would contribute to enabling successful execution of the subsequent FDI and sensitivity analysis.

Independent variables

Independent variables in the context of the TEP simulation model for this study would serve to direct the nature of operational conditions experienced during simulation. This would in turn serve to generate multiple datasets that are consistent in representing each respective fault condition, while demonstrating sufficient variation in the data to be used for FDI and subsequent sensitivity analysis. The following independent variables, to influence operational conditions, were identified that would suit this purpose:

- Duration of simulated operational conditions (NOC & 20 Faults)
- Duration of simulated reference conditions (NOC & 20 Faults)
- Duration of normal operation before introducing fault (Operational data)
- Duration of normal operation before introducing fault (Reference data)
- Size of disturbance leading to fault condition (Per fault condition)
- Control scheme selection
- Mode selection (Modes 1 - 6)
- Randomness seeds implemented

Dependent variables

Following the identification of independent variables of the TEP simulation model in the context of this study, the dependent variables indicate the resulting changes observed in the operational condition datasets. Variation in the following variables, as properties of the operational condition data, would prove to be beneficial towards the FDI and sensitivity analysis of functional parameters to be performed for this study:

- Ratio of NOC to fault condition data samples (per simulation).
- Variability of individual faults between datasets.
- Quantity of fault condition samples in operation and reference data.
- Variation in each fault condition's initial disturbance.

Change to the above dependent variables will not be measured individually, as this would require analysis of process data acquired to verify possible changes and would exceed the scope of this study. Instead, the change to the above dependent variables would be initially assumed as correct towards motivation in acquiring multiple simulated datasets containing sufficient variation. Change to the above variables would instead be observed through deviations in performance measures from the FDI results obtained.

4.2.2 Simulation model

Process measurements are freely accessible within the MATLAB workspace, allowing interaction at the user's discretion, and are easily exported to external applications. Figure 4.1 illustrates the high level model hierarchy of the Simulink model by [12], where properties of individual feed streams and process controllers can be accessed to change operating conditions of the model. This includes the opportunity to set initial and setpoint values to replicate the 6 predefined operating modes presented in [7]. Selecting any of the sub-system blocks would present further in-depth detail. Figure 4.2 presents the Condenser sub-system block and the underlying processes, with respective process input and output variables associated with its operation. These respective process blocks could be used for further navigation to lower-level components or processes within the system.

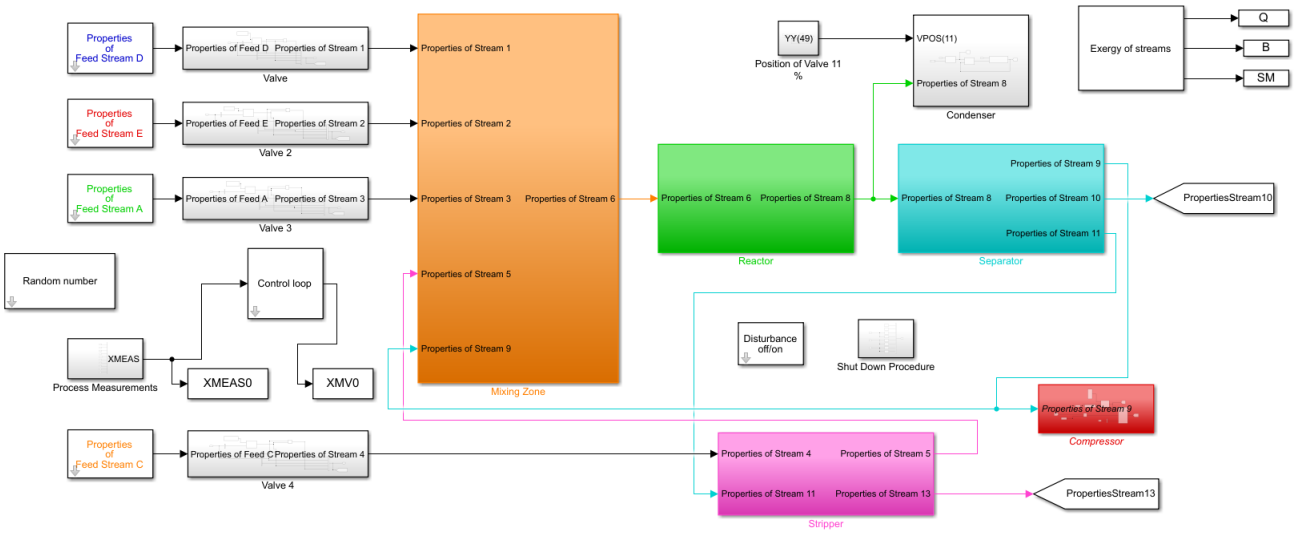

Figure 4.1: Simulink structure of the TEP model

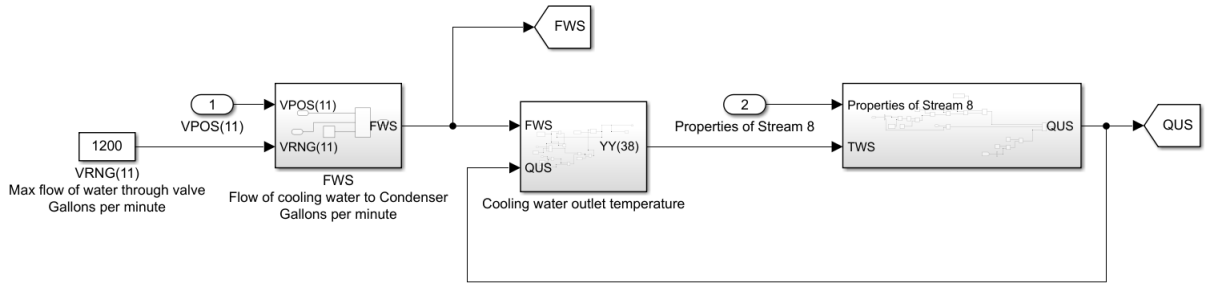

Figure 4.2: Condenser sub-system block in the TEP model

Management of system parameters in the TEP model is accessible for user preference to influence the system operating conditions as required. The process parameters made available for influencing the TEP model's operation are material stream properties, process setpoints, and controller setpoints implemented towards maintaining process operations within system constraints. The initial TEP model was based on an existing production facility connected to other upstream- and downstream process facilities where production could change based on product demand. Depending on the product selected, initial starting conditions would change according to the initial composition of reactants in the four feed streams supplied from other processes within the plant facility. Similar to industry conditions [7] notes "the plant production rate is set by market demand or capacity limitations" where these production rates are set by the mode selected [7]. Altering system operation is available through selection of the control scheme per user preference and by the manipulated variable setpoints to influence system operation conditions. These changes directly alter the constraints by which sub-process components are required to maintain acceptable operating conditions. The control strategy to be implemented can be selected from one of the following:

- The initial open loop scheme presented by [7]
- Closed loop control based on one of the structures presented in [11] as implemented by [8].
- Control scheme demonstrated by [9] implementing decentralised control by means of multiple independent controllers.

Initial process measurement data obtained by executing the TEP simulation model for dataset 5, fault are presented in Figure 4.3, where a section of initial process values for process measurements (XMEAS) 6, 7, 8, 9 and 21, which are associated with the reactor sub-system in the TEP model, are illustrated. Figure 4.4 similarly presents dataset 5, fault 6 process values for XMEAS 37, 38, 39, 40 and 41 which are associated with the stripper responsible for removing any reactant material present in the product material stream.

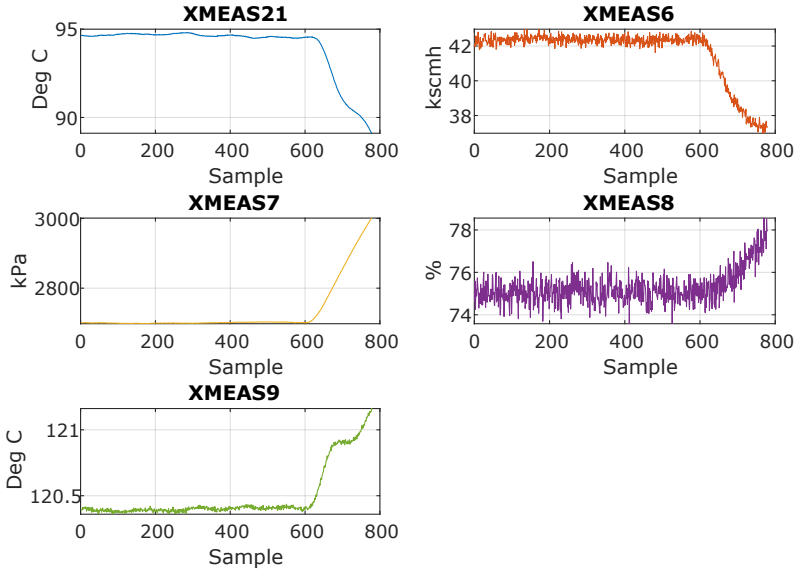

Figure 4.3: TEP reactor measurements - Dataset 5 Fault 6

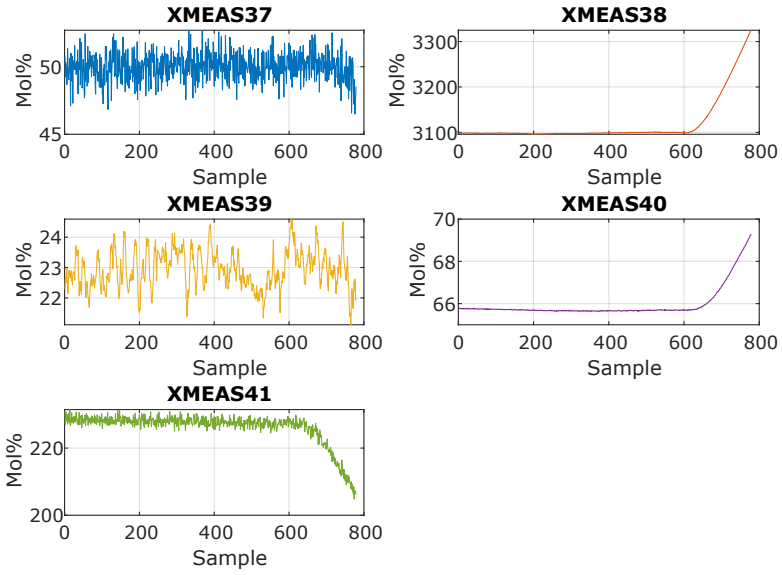

Figure 4.4: TEP stripper measurements - Dataset 5 Fault 6

4.2.3 Data acquisition

The TEP simulation model is configured to perform 21 simulation executions, representing the NOC and each of the 20 fault conditions, and compiles this data as a single dataset. This dataset would then contain data sampled at the various rates of the individual process instrumentation recreated in the model, and would need to be reduced to a single sample rate similar to [8]. This will be further discussed in the following section.

Multiple datasets which are to be generated in this study using the above-mentioned independent variables would only include single variable variation of those previously identified. Alteration of too many variables would introduce uncertainty and require verification of the data to be performed before it is considered valid for use. The intended variation of the randomness seeds implemented, as independent variable, would allow variation of operational conditions while not altering system operation setpoints or the initial disturbance leading up to fault conditions experienced. The variables responsible for simulation duration and size of disturbance were kept constant to ensure sufficient data samples and disturbances remained consistent per fault condition across all datasets simulated. A quantity of 10 datasets were selected in providing a sufficient quantity of datapoints towards the sensitivity analysis, compared to the cost in time required to execute these simulations. Variation to the randomness seed would differ between each of the 21 operational conditions per operation and reference data simulated. Further variation of this would also be controlled between datasets to prevent duplication of operational condition data when using identical values. Information on these random variation seeds can be found under Appendix B, where the range of values available for use and order implemented per dataset can be viewed.

The duration of simulated operational conditions, i.e. both operation and reference data, will be kept constant to the values used by [12] and [8] as these have previously been proven successful while serving as known reference for expected fault conditions. This would contribute towards establishing whether the FDI methods performed their intended purpose successfully when presented with a large quantity of operation data to discern the respective reference conditions. The following duration values would be implemented for each of the datasets simulated for this study:

- Duration of simulated reference condition: 25 Hours
- Duration of simulated operation condition: 48 Hours
- Duration of normal operation prior to reference condition start: 1 Hour
- Duration of normal operation prior to operation condition start: 8 Hours

The size of disturbances as cause to fault conditions experienced will be kept constant at the default value (100%), as changing this variable would not constitute a valid assumption to this change resulting in direct change to fault conditions observed, without first performing verification prior to use in this study. Exploring FDI and the subsequent sensitivity analysis across an operational envelope is not included as part of the scope for this study.

The control mode will be kept constant with the "closed-loop control system" selected for all simulations performed. Mode selection will be utilised towards providing separate operational conditions of the TEP model to allow validation, using data from the same system. Initial datasets would be generated with the TEP model operating in the "Base case mode" or "Mode 1" (depending on literature viewed), while datasets for validation would be simulated at an alternate mode. Doing so would allow validation based on like-for-like comparison between FDI and sensitivity analysis results.

4.3 FDI data utilisation

4.3.1 Transformation of process measurement data

All simulated data generated, as operation and reference data separately, would first be transformed to implement a consistent sampling rate amongst process measurement data. Sampling intervals differ amongst the process measurements ranging from 3 seconds, 360 seconds and 900 seconds whereas many of the time constants of the TEP under closed loop control are close to 2 hours [8]. This study did not implement a low pass filter for data sampled and instead sampled at half the interval of the slowest feed stream process measurements. The reactant material feed leading to the reactor has a slow sampling interval of 360 seconds where best practice is to sample at half this rate to prevent aliasing of the signal observed. This consisted of sampling at 180 seconds and was used as the base rate to which all other process measurement data was transformed to [8]. This reduced sample rate is demonstrated in Figure 4.5 below, where the number of samples in both operation and reference data is reduced by a factor of 180, leading to a significant reduction in the process sample count in each group. Russel et al. presents a discussion on considering using moving average filter to reduce process variability at the increased computational cost and delaying fault detection. Russel et al. further reaffirms this method of using every 180th sample is suitable for fault detection purposes [8].

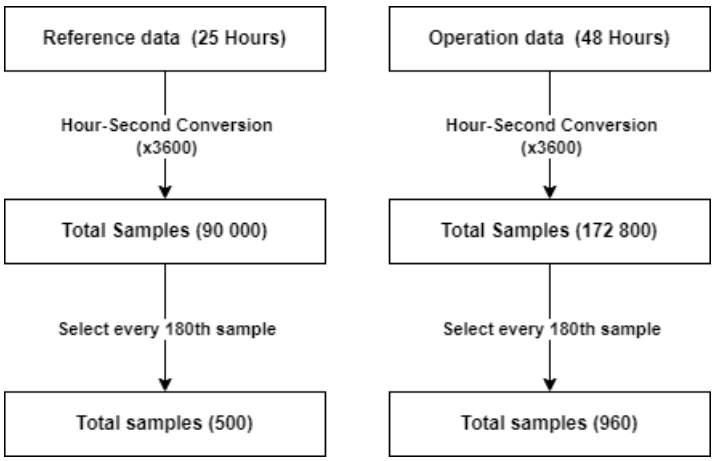

Figure 4.5: Conversion of TEP model sampling rate

The transformation of raw process measurements into data fit for use, containing energy and exergy flow attributes, follows the same method demonstrated by [12] where exergy-based FDI was performed using the TEP model. This includes using the same notation for TEP model components and calculations compiled in MATLAB [12]. A demonstration of this notation can be found in Figure 4.6, where the TEP model is represented by nodes with interconnected lines indicating interaction between respective nodes. This notation would then be used to compile attributed graphs, where nodes will determine matrix position and be filled with the corresponding energy- and exergy-flow data between nodes.

Reference graph

Each operational condition, either normal or fault condition, will be represented by a single reference graph serving as known truth, by which operational conditions would be compared to when performing FDI. Each reference graph was constructed from the average value of the 500 operational data samples, with a simulation duration of 24 hours for each position in the node signature matrix.

Operation graph

The operation data delivered by the TEP model demonstrates either normal conditions or one of the respective fault conditions experienced during model operation. These represent the operational conditions to be detected and isolated when performing FDI using the respective EGBV methods. Similar to the reference graph, operational data will be packaged per attributed graph notation in (7) to create the operation graphs to be used when performing

FDI. These operation graphs then represent the operational condition to be identified by the respective EGBV methods when performing FDI. The database of operation graphs would consist of 21 series of graphs, for the 21 operational conditions simulated per dataset, with graphs being kept in their respective order per operational condition [6].

4.3.2 FDI using functional parameter permutations

Following identification of functional parameters in Chapter 3, FDI will be performed using combinations of these functional parameters for the respective EGBV methods. Each combination per EGBV method, previously displayed in Figure 3.23, would be executed using the same series of datasets to ensure a valid comparison of FDI results.

The functional parameter combinations, referred to as permutations going forward, would be executed independently with FDI results being compared side-by-side once completed. Side-by-side comparison of FDI results by these permutations would serve to illustrate the performance of the default permutation, as well as provide more discernible indications to deviation by the other permutations. Figure 4.7 demonstrates an example of the table format to be used when comparing FDI results by permutations for the Distance approach. A similar format is implemented when discussing FDI results per the individual performance metrics for the other EGBV methods in the following chapter.

	HEOM MEAN	HEOM RC: SMALLEST 1	HEOM RC: SMALLEST 5	HEOM RC: SMALLEST 10	HEOM Threshold: 2-SD MIN	HEOM Threshold: 3-SD MIN
ACCURACY						
ISOLATION						
MCC						

	SHEOM MEAN	SHEOM RC: SMALLEST 1	SHEOM RC: SMALLEST 5	SHEOM RC: SMALLEST 10	SHEOM Threshold: 2-SD MIN	SHEOM Threshold: 3-SD MIN
ACCURACY						
ISOLATION						
MCC						

	IQR-HEOM MEAN	IQR-HEOM RC: SMALLEST 1	IQR-HEOM RC: SMALLEST 5	IQR-HEOM RC: SMALLEST 10	IQR-HEOM Threshold: 2-SD MIN	IQR-HEOM Threshold: 3-SD MIN
ACCURACY						
ISOLATION						
MCC						

Figure 4.6: Example of FDI result table: Distance approach

4.3.3 Performance metrics

The performance measurements to be used when evaluating FDI results delivered by the EGBV methods in this study are listed below, and are accompanied by a brief description explaining their purpose in the context of this study. FDI results delivered by the permutations are then inspected side-by-side, presented in the previous section, to illustrate possible deviation when compared to the default permutation per EGBV method [19].

Accuracy	Rate of correct detection classifications compared to the total range of classifications [26].
Isolation	Determine the location and kind of fault (after fault detection) and measure these classifications per the range of faults considered [34].
MCC	Geometric mean of the four confusion matrix components [35].

The implementation of MCC assigns equal weight to both true negative and true positive classifications, as opposed to other measures only considering true positive classifications during their calculation. A conservative approach was taken to assume that any change in functional parameter for an FDI method immediately alters its ability to clearly identify normal conditions. This approach served to identify cases where an FDI method permutation would classify all conditions observed as a fault. Such cases could present a relatively good accuracy as false indicator to the detection capability by the respective FDI method permutation. Implementation of the MCC as performance meaure does not directly affect the FDI method's ability to discern between fault and normal conditions. Instead, it is an indicator on whether changes to functional parameters altered the FDI method's ability to correctly discern the normal condition from all fault conditions presented. Conventional use of accuracy and isolation as performance measures do not take discerning valid normal operation conditions into account for their calculation. The incorrect classification of true negative conditions could create the potential for uncertainty in whether the presented true positive classifications are valid, or if the FDI method simply detected all conditions as fault conditions. Therefore, awarding equal weight to a single true negative versus the multiple true positive classifications makes the MCC well suited as metric for evaluating FDI performance [35]. Table 4.1 provides clarification on the range, unit, and interpretation of values expected for each of these performance measures.

Table 4.1: Performance measure clarification

Measure	Range	Unit	Interpretation
Accuracy	[0 , 100]	%	Worst (0), Best(100)
Isolation	[0 , 100]	%	Worst (0), Best(100)
MCC	[-1 , 1]	Unitless	Worst(-1), Bad(≤ 0), Good(> 0), Best(1)

Based on the confusion matrix on detection classifications for fault conditions observed by an FDI method [5, 6], Table 4.2 presents the key to both accuracy and isolation classifications of fault conditions observed by the FDI methods. In the event of both a true and negative detection classification being detected during evaluation of an operational condition, preference would be given to the negative classification. Such a case in performing isolation would instead be classified as unisolated for the observed operation condition.

Table 4.2: FDI classifications

Result	Context
Detection	
TrueNeg	True negative
FalsePos	False positive
FalseNeg	False negative
TruePos	True positive
Isolation	
Missed	Missed isolation
Correct	Correct isolation
Unisolated	Unable to isolate condition

4.4 Conclusion

The initial form of process data delivered as model output through the use of the TEP simulation model was discussed, followed by a discussion on how this data would be further transformed into representing energy and exergy-flow present within the system during operation conditions experienced.

Following this data transformation, the structure by which energy and exergy-flow data would be packaged as attributed graphs was discussed, where initial operation and reference data generated would then be utilised as operation- and reference graphs when performing FDI. FDI using the EGBV methods, as permutations of functional parameters, will then be performed in the following chapter, where the initial discussion is presented on how the FDI results would contribute towards the sensitivity analysis of functional parameters. FDI and sensitivity analysis of functional parameters is then performed per EGBV method, where the results delivered by the respective permutations are compared.

Chapter 5

Sensitivity analysis of EGBV functional parameters

5.1 Introduction

Following data compilation for datasets acquired, FDI is then performed using the compiled attributed graphs containing energy and exergy-flow data during operational conditions experienced by the TEP model. Comparing FDI results delivered by EGBV method permutations would present an opportunity to observe the initial deviation from a respective method's default permutation. Performance by these EGBV method permutations would then serve as input to the sensitivity analysis of functional parameters performed in this chapter, to address the third research objective identified for this study.

Execution of the FDI will be structured in performing all permutations per EGBV method where FDI results could be inspected concurrently. This comparison includes inspection of the accuracy, isolation and MCC performance measured. The sensitivity analysis would then utilise these FDI results delivered by EGBV method permutations, to identify potential sensitivity demonstrated through variation of the functional parameters implemented.

5.2 FDI results for ANOVA

Analysis of variance (ANOVA) as sensitivity analysis method is suited for the application in this study, as the functional parameters are qualitative. Variation to their value is based on a property instead of a numerical value, discrete or continuous, that could be analysed within a range of possible values. The FDI results delivered through the use of multiple datasets only containing variation to the randomness seeds, present an opportunity to perform two-way ANOVA with replication. This form of ANOVA would analyse variation of two variables in the input data to determine sensitivity towards each based on the variation in the

repeated measures. These repeated measures are based on performing multiple experiments or observations without system variation to establish variance amongst consistent/constant system conditions.

Purpose of performing ANOVA

The implementation of two-way ANOVA with repeated measures aims to identify possible sensitivity demonstrated by the EGBV methods towards the respective functional parameters. The proposed sensitivity analysis serves to confirm the presence of sensitivity or lack thereof towards individual functional parameters identified through an analysis of the accuracy, isolation and MCC values delivered by the EGBV method permutations. Sensitivity is then evaluated individually by each performance measure towards each of the functional parameters implemented.

ANOVA and functional parameter sensitivity

Method: Performing a two-way analysis of variance (ANOVA) on two independent variables to determine the significance of their influence on the considered output variable by means of statistical inspection of overlapping indices. This method is beneficial in allowing the analysis of categorical variables, such as fault type, and quantitative variables simultaneously when given the correct number of intervals to ensure sufficient resolution of table entries. Consider the following methodology used in performing two-way ANOVA with replication [30]:

Define the null and alternative hypotheses for each of the considered independent variables. The null hypothesis states that there is no main effect of an independent variable on the response observed. The alternative hypothesis stands as an opposing hypothesis, where a main effect is demonstrated by the independent variable on the response observed [36]. An example is demonstrated below:

$$H_0 : \text{There is no significant difference in group means.}$$
$$H_a : \text{There is significant difference in group means.}$$

Select a significance threshold level α by which interaction levels would be measured. Interaction levels above the α disproves the null hypothesis and rather validate the alternative. A threshold value of $\alpha = 0.05$ is commonly used.

Calculate the total sum of squares for all observations with the correction for the mean (CM). Equation (25) demonstrates the total sums of squares for an ANOVA table, with r rows and c columns based on the size of options (i, j) per independent variable. CM indicates the mean

correction of total observations B by number of samples n as demonstrated in (26) [30].

$$SS_{Total} = \sum_{i=1}^{r}\sum_{j=1}^{c}(x_{ij})^2 - CM \tag{25}$$

$$CM = \frac{(B)^2}{c \times r \times n} \tag{26}$$

Next, the sum of squares is calculated for each group representing the respective independent parameters to be analysed. The column-wise group is presented in (27) as comprised of the square sum of column totals B_j minus the mean correction for all samples [30].

$$SS_{Column} = \frac{\sum_{j=1}^{c} {B_j}^2}{c \times n} - CM \tag{27}$$

The sum of squares is calculated similarly for the row-wise group seen in (28), for the square sum of row totals B_i [30].

$$SS_{Row} = \frac{\sum_{i=1}^{r} {B_j}^2}{r \times n} - CM \tag{28}$$

Interaction between row and column groups is calculated using the above sum of squares. Equation (29) demonstrates that this interaction consists of the sum of squared totals for each table index, row-wise and column-wise totals B_{ij}, divided by the number of repetitions n. This total is then reduced with both the row and column sum of squares, and the mean correction to deliver interaction observed between the two independent variables analysed [30].

$$SS_{Interaction} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c} {B_{ij}}^2}{n} - \frac{\sum_{j=1}^{c} {B_j}^2}{c \times n} - \frac{\sum_{i=1}^{r} {B_i}^2}{r \times n} + CM \tag{29}$$

The error within sample squares in (30) is calculated as the difference between the total SS_{Total} and the sum of row, column and interaction squares. This demonstrates the variation observed amongst samples that could not be attributed to the respective independent variables analysed [30].

$$SS_{Within} = SS_{Total} - SS_{Row} - SS_{Within} - SS_{Interaction} \tag{30}$$

The degree of freedom for each of these sum of squares is calculated based on the groups involved, as demonstrated in Table 5.1 [30] below. The Mean Square and resulting F-counted values are then calculated for each of the respective sources of variation identified.

Table 5.1: Example of two-way ANOVA

Source of Variation	**df**	**Mean Square**	**F-Value**
Row	$d_{Row} = r - 1$	$MS_{Row} = \frac{SS_{Row}}{df_{Row}}$	$F_{Row} = \frac{MS_{Row}}{MS_{Error}}$
Column	$d_{Column} = c - 1$	$MS_{Column} = \frac{SS_{Column}}{df_{Coluumn}}$	$F_{Column} = \frac{MS_{Column}}{MS_{Error}}$
Interaction	$d_{RC} = (r-1)(c-1)$	$MS_{RC} = \frac{SS_{RC}}{df_{RC}}$	$F_{RC} = \frac{MS_{RC}}{MS_{Error}}$
Error	$d_{Error} = (r \times c)(n-1)$	$MS_{Error} = \frac{SS_{Error}}{df_{Error}}$	
Total	$df_{Total} = (r \times c \times n) - 1$		

The sums of squares are utilised for calculating the F-values in order to compare it to the Fisher distribution tables (F-critical values) to determine the statistical significance of the variation observed. Comparison of the F-values to such a table would deliver the following classifications:

- If F-value is less than F-critical value, the null hypothesis can be accepted as no significant cause to variance is observed.

- If F-value is equal or greater than F-critical value, the null hypothesis can be rejected and the alternate hypothesis could instead be accepted as significant cause to variance was observed.

Following ANOVA performed, the sensitivity demonstrated towards each of the respective sources of variation is calculated as the share of the total sums of squares observed for the analysis. Equation (31) in calculating sensitivity is accompanied by examples of how sensitivity towards the normalisation (row), cost-matrix utilisation (column), interaction between parameters, and variation within sample groups will be calculated in this chapter [37].

$$Sensitivity = \frac{SS_{Source}}{SS_{Total}} \tag{31}$$

$$Sensitivity_{(Normalisation)} = \frac{SS_{Normalisation}}{SS_{Total}}$$

$$Sensitivity_{(Cost-Matrix\ Utilisation)} = \frac{SS_{Cost-Matrix\ Utilisation}}{SS_{Total}}$$

$$Sensitivity_{(Interaction)} = \frac{SS_{Interaction}}{SS_{Total}}$$

$$Sensitivity_{(Within)} = \frac{SS_{Within}}{SS_{Total}}$$

5.3 Distance approach

The Distance approach presented in Chapter 3 to identify applicable functional parameters for this method was further expanded in performing FDI, by utilising the operation graph as a series to represent the operational condition to identify. This includes observing the distance calculated between operation- and reference graphs as the sum of classifications allocated with each respective reference condition. Classifications allocated to each reference condition would then be represented as a percentage of total classifications delivered. From the simulation configuration previously discussed in the previous chapter, the operation graphs per dataset would consist of the following:

- The NOC consists of 960 operation graphs (160 for initial 8 hours, 800 for remaining 40 hours simulated).
- Each fault condition consists of 800 graphs per the 40 hours simulated after initial normal operation.

The Distance approach was executed as per the range of functional parameter permutations by using the acquired datasets in the attributed graph format, as previously demonstrated in Chapter 4 above. This section discusses the FDI results obtained for these functional parameter permutations according to the identified performance measures, and followed by performing ANOVA using these values in order to determine whether significant sensitivity is demonstrated.

5.3.1 FDI: Distance approach

All permutations for the Distance approach were applied concurrently to each of the respective datasets to establish their respective performances in detecting and isolating operational conditions observed. Each permutation attempts to detect and isolate conditions observed according to the reference graphs presented in each dataset. Table 5.2 demonstrates the FDI results by the HEOM-Mean (default permutation), where the series of operation graphs per condition (Left) is compared to the respective reference graph database (Top) from dataset 2. The detection and isolation classifications delivered by this permutation seen on the right-hand side of the figure present a fair sense of the capability in delivering FDI classifications. The values demonstrated indicate the percentage of classifications assigned to the reference graph (column), to which the operation graphs (row) demonstrated the highest similarity as viewed by the default permutation. The main diagonal of the table indicates where reference and operation conditions correspond, and where correct isolation classification for a condition would be located in the table. From the results seen in Table 5.2, each condition observed a degree of similarity with the NOC, where cases of false negative classifications further demonstrated the majority share of operation graph similarity that has been assigned to the NOC.

Similar FDI performance was demonstrated by the SHEOM RC: SMALLEST 10 permutation where a reduced capacity of the confusion matrix was utilised. Results seen in Table 5.3 demonstrate that this permutation is able to correctly detect conditions that were missed in comparison to the default permutation in Table 5.2, while also demonstrating false negative classifications that were previously correctly detected by the default permutation. A corresponding degree of similarity between operation graphs and the NOC reference graph is observed in this case.

Notable in these results is the case to the step fault type, demonstrating reduced quantity/share of incorrect classifications when comparing similarity between operation and reference graphs. In contrast, locations to incorrect classifications are more spread across the remaining fault types.

Table 5.2: FDI Results: HEOM-MEAN - Dataset 2

		REFERENCE GRAPH DATABASE																				FDI OUTCOME	
		NOC	FAULT 1	FAULT 2	FAULT 3	FAULT 4	FAULT 5	FAULT 6	FAULT 7	FAULT 8	FAULT 9	FAULT 10	FAULT 11	FAULT 12	FAULT 13	FAULT 14	FAULT 15	FAULT 16	FAULT 17	FAULT 18	FAULT 19	FAULT 20	DETECTION	ISOLATION
OPERATION GRAPHS	NOC	95.42	0	0	0	0	0	0	0	0	0	0.31	0.21	0.21	0	0.10	0	0.21	0.10	2.71	0	0.73	TrueNeg	Correct
	FAULT 1	16.67	83.02	0.21	0	0	0	0	0	0.10	0	0	0	0	0	0	0	0	0	0	0	0	TruePos	Correct
	FAULT 2	16.67	0	83.02	0	0	0	0	0	0	0	0	0.10	0	0.10	0.10	0	0	0	0	0	0	TruePos	Correct
	FAULT 3	17.50	0	0	81.35	0	0	0	0	0	0.73	0	0	0	0	0	0	0	0	0.21	0	0.21	TruePos	Correct
	FAULT 4	89.90	0	0	0	0.10	0	0	0	0.21	0.10	0.42	0	0.21	4.69	0	0.94	0	0.21	2.50	0	0.73	FalseNeg	Missed
	FAULT 5	16.67	0	0	0	0	80.00	0	0	0	0	0	0	3.33	0	0	0	0	0	0	0	0	TruePos	Correct
	FAULT 6	16.67	77.29	0.73	0	0	0	0	0	0	5.21	0	0	0	0	0	0	0.10	0	0	0	0	TruePos	Missed
	FAULT 7	16.67	0	0.73	0	0	0	0	0	0.21	0	0	0	0	0	81.88	0	0.10	0	0	0.42	0	TruePos	Missed
	FAULT 8	16.67	0	6.04	0	0	0	0	0	0.10	0	0.10	0	0	76.98	0	0	0	0	0	0	0.10	TruePos	Missed
	FAULT 9	92.08	0	0	0.10	0	0	0	0	0	0.52	0	0.10	0	0	0.42	5.94	0.42	0	0.42	0	0	FalseNeg	Missed
	FAULT 10	16.67	0	0.10	0	0	0	0	0	0	5.10	0.42	0	0	0	0.63	66.04	0.10	0	0	10.94	0	TruePos	Missed
	FAULT 11	16.67	0	0	0	0.10	0	0	0	0	0.31	0	60.21	0	0	21.77	0	0	0	0	0	0.94	TruePos	Correct
	FAULT 12	83.96	0	7.50	0	0	0	0	0	0.21	0	0.42	0	0	0	0.31	4.58	0.10	0	1.88	0	1.04	FalseNeg	Missed
	FAULT 13	16.67	0	0	0	0	0	0	0	0.10	9.58	0.52	0	0	0	0.73	61.15	0	0	0	11.25	0	TruePos	Missed
	FAULT 14	16.67	0	0	0	0	0	0	0	0	0	0.10	0	0	0	82.50	0	0.10	0	0	0.63	0	TruePos	Correct
	FAULT 15	93.23	0	0	0	0	0	0	0	0	0	0.10	0	0	1.25	0	1.77	0	0.10	3.23	0	0.31	FalseNeg	Missed
	FAULT 16	48.85	0	46.67	0	0	0	0	0	0.10	0	0.31	0	0	0.31	0.10	0	0.42	0	2.60	0	0.63	FalseNeg	Missed
	FAULT 17	35.00	0	0	0	0	0	0	0	0.10	0.42	0.21	0.94	0	0	58.23	0.52	0.10	0	0.94	0	3.54	TruePos	Missed
	FAULT 18	16.67	0	0	0	0	0	0	0	0.42	0.10	0.21	0	0.10	0	6.25	0.21	0	0.10	75.73	0.21	0	TruePos	Correct
	FAULT 19	16.67	0	0	0	0	0	0	0	0.10	0	0.42	0	0	0	0.31	0.21	0	0	0	82.29	0	TruePos	Correct
	FAULT 20	42.81	0	0	0	0	0	0	0	0	0	0.10	0	0	0	1.35	0.42	0	0	1.46	0.94	52.92	TruePos	Correct

Table 5.3: FDI Results: SHEOM RC: SMALLEST 10 - Dataset 6

		REFERENCE GRAPH DATABASE																				FDI OUTCOME	
		NOC	FAULT 1	FAULT 2	FAULT 3	FAULT 4	FAULT 5	FAULT 6	FAULT 7	FAULT 8	FAULT 9	FAULT 10	FAULT 11	FAULT 12	FAULT 13	FAULT 14	FAULT 15	FAULT 16	FAULT 17	FAULT 18	FAULT 19	FAULT 20	DETECTION	ISOLATION
OPERATION GRAPHS	NOC	95.10	0	0	0	0	0.10	0	0	0.21	0.10	0	0	0	0.21	0	0.21	0.21	2.81	0.52	0	0.52	TrueNeg	Correct
	FAULT 1	16.67	64.38	0.10	0	0	0	0	0	18.54	0	0	0.21	0	0	0	0.10	0	0	0	0	0	TruePos	Correct
	FAULT 2	17.60	0	81.56	0	0	0	0	0	0.31	0	0.21	0	0	0	0	0	0	0	0.21	0	0.10	TruePos	Correct
	FAULT 3	24.79	0	0	71.46	0	0	0	0	0	1.35	0	0	0.10	0	0	0.52	0	1.56	0.21	0	0	TruePos	Correct
	FAULT 4	16.67	0	0	0	2.50	0	0	0	0	0	0	0.10	0	0	0	0	0	76.35	0	0	4.38	TruePos	Missed
	FAULT 5	16.67	0	0	0	0	83.13	0	0	0	0	0	0	0.10	0	0	0	0	0	0	0	0.10	TruePos	Correct
	FAULT 6	16.67	0.94	0.31	0	0	0	0.21	0	81.35	0	0	0	0	0.21	0	0	0	0	0	0.21	0.10	TruePos	Missed
	FAULT 7	16.67	0	0	0	0	0	0.31	63.85	0	0	0.42	0	0.10	0	0	0	0.10	0	17.29	0	1.25	TruePos	Correct
	FAULT 8	63.54	0	30.42	0	0	0	0	0	0	0.31	0.31	0.10	0	0.31	0	1.98	0.10	2.08	0.21	0.10	0.52	FalseNeg	Missed
	FAULT 9	27.19	0	0	0	0	0	0	0	0.10	10.63	0	0.42	0	1.46	0.21	57.29	0.31	2.08	0	0	0.31	TruePos	Missed
	FAULT 10	35.83	0.10	0	0	0	0	0	0	0	3.75	9.79	0	0.21	3.65	0	44.06	0.42	1.46	0.21	0	0.52	TruePos	Missed
	FAULT 11	92.92	0	0	0	0	0	0	0	0	0.42	0.10	0.21	0.10	0.52	0.21	1.67	0.31	2.50	0.52	0.31	0.21	FalseNeg	Missed
	FAULT 12	18.44	0	0	0	0	0	0	0	0	10.10	14.17	0	0	0.63	0.31	52.60	0.31	2.92	0.10	0	0.42	TruePos	Missed
	FAULT 13	71.35	0	0	0	0	0	0	0	0	0	0.21	0	0	3.96	0.10	22.81	0	0.83	0.42	0	0.31	FalseNeg	Missed
	FAULT 14	16.77	0	0	0	0	0	0	0	0	0	0	0	0	0	83.02	0	0.10	0	0	0	0.10	TruePos	Correct
	FAULT 15	74.27	0	0	0	0	0	0	0	0	0.94	0	0.10	0	0.31	0.21	20.42	0	2.81	0	0	0.94	FalseNeg	Missed
	FAULT 16	80.31	0	0.10	0	0	0	0	0	0.10	0.42	0.10	0.10	0	0.42	0	0.73	15.00	2.71	0	0	0	FalseNeg	Missed
	FAULT 17	16.67	0	0	4.27	0.21	0	0	0	0	0.31	0	0	0	5.94	0.10	0	0	8.65	0	63.85	0	TruePos	Missed
	FAULT 18	16.67	0	0	0	0	0	0	0	0	0	0	0	0.83	0	0.10	0.10	0.21	0	80.31	1.67	0.10	TruePos	Correct
	FAULT 19	81.98	0	0	0	0	0	0	0	0	0.63	0	0	0.10	0.31	0	12.60	0.42	3.44	0.31	0	0.21	FalseNeg	Missed
	FAULT 20	18.44	0	0	6.15	0	0	0	0	0.10	2.50	0	0	0	1.35	12.29	51.15	0.31	0.83	0.10	0	6.77	TruePos	Missed

Table 5.4: FDI Results: IQR-HEOM Threshold: 2-SD MIN - Dataset 9

		REFERENCE GRAPH DATABASE																				FDI OUTCOME	
		NOC	FAULT 1	FAULT 2	FAULT 3	FAULT 4	FAULT 5	FAULT 6	FAULT 7	FAULT 8	FAULT 9	FAULT 10	FAULT 11	FAULT 12	FAULT 13	FAULT 14	FAULT 15	FAULT 16	FAULT 17	FAULT 18	FAULT 19	FAULT 20	DETECTION	ISOLATION
OPERATION GRAPHS	NOC	89.58	0	0	0	0	0	0	0	0.10	0.10	0.10	0	0	0.52	0	6.35	0	0.52	1.46	0	1.25	TrueNeg	Correct
	FAULT 1	16.77	62.40	0.10	0	0	0	0	0	20.00	0	0	0	0	0.21	0	0.42	0	0	0.10	0	0	TruePos	Correct
	FAULT 2	16.77	0	82.60	0	0	0	0	0	0	0	0	0	0	0.10	0.21	0.31	0	0	0	0	0	TruePos	Correct
	FAULT 3	20.83	0	0	72.19	0	0	0	0	0	0.63	0	0	0.21	0.21	0	2.81	0	0	2.08	0	1.04	TruePos	Correct
	FAULT 4	16.77	72.50	0	0	0.10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.31	10.31	TruePos	Missed
	FAULT 5	16.67	0	0	0.10	0	68.02	0	0	0	0	0	0	15.21	0	0	0	0	0	0	0	0	TruePos	Correct
	FAULT 6	16.67	0.31	0.63	0	0	0	0	0	20.21	0	0	0	0	0.10	0.63	0	0.10	0.21	0.10	57.92	3.13	TruePos	Missed
	FAULT 7	16.67	5.21	0	0	0	0	0	0	0.52	0	2.19	0	0	0	70.31	0	0	0	0.31	0.21	4.58	TruePos	Missed
	FAULT 8	48.02	0	12.81	0	0	0	0	0	29.48	0.31	0.10	0.10	0	0.83	0	5.63	0	0.63	1.25	0	0.83	FalseNeg	Missed
	FAULT 9	16.67	0	0	82.60	0	0	0	0	0	0.63	0	0	0.10	0	0	0	0	0	0	0	0	TruePos	Missed
	FAULT 10	73.44	0	0.10	0	0	0	0	0	0	0.63	0.10	0	0.10	0.52	0.10	22.19	0.10	0.10	1.35	0.10	1.15	FalseNeg	Missed
	FAULT 11	16.67	0	0	0	83.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	TruePos	Missed
	FAULT 12	30.73	0	0.31	0	0	0	0	0	23.65	7.19	0.21	0	0.10	3.02	0.63	33.02	0	0	0.42	0.10	0.63	TruePos	Missed
	FAULT 13	33.54	0	61.15	0	0	0	0	0	0	0	0.21	0	0	0.42	0.10	3.13	0	0.10	0.52	0	0.83	TruePos	Missed
	FAULT 14	31.67	0	0	0	0	0	0	0	0.10	0	0.10	0	0	0.10	0.10	10.10	0.10	0	0.42	56.88	0.42	TruePos	Missed
	FAULT 15	87.19	0	0	0	0	0	0	0	0	0	0.10	0	0	0.63	0	8.85	0	0	1.77	0	1.46	FalseNeg	Missed
	FAULT 16	67.29	0	0	0	0	0	0	0	11.15	0	0.31	0	0	6.98	0	8.54	0.10	0	1.25	0	4.38	FalseNeg	Missed
	FAULT 17	16.77	3.54	0.10	0	0	0	0	0	0.52	0	0.10	0	0	0.10	0.10	0	0	0	0	1.56	77.19	TruePos	Missed
	FAULT 18	16.67	0	0	0.73	0	0.10	0	0	3.44	0	0.21	0	34.79	2.60	3.54	2.19	0	0.10	35.52	0	0.10	TruePos	Correct
	FAULT 19	16.67	0	0	0	0	0	0	0	0.10	0	0	0	0	0.21	0	0.63	0	0.31	0.10	81.88	0.10	TruePos	Correct
	FAULT 20	16.67	0	0	0	0	0	0	0	0.10	0	0.21	0	0	0.10	0.10	12.92	0	0	0.10	17.19	52.60	TruePos	Correct

The IQR-HEOM Threshold: 2-SD MIN demonstrated a moderately high rate to correct fault detection, with the added benefit of correctly delivering a true negative classification when observing the NOC operation for dataset 9. The same observation to reduced quantity of incorrect assignments by the Distance approach can be seen in Table 5.4, which contains the previously mentioned results. Results illustrated in these figures only represent a brief

indication of the complete FDI performed by utilising the 18 Distance approach permutations applied across 10 datasets.

FDI results delivered by the Distance approach permutations when applied to the 10 datasets were compiled to deliver a mean value for accuracy, isolation, and MCC respectively. Table 5.5 demonstrates these mean values for each of the 18 permutations executed. The HEOM-MEAN as default permutation demonstrated fair accuracy in correctly detecting operational conditions experienced, while the majority of alternate permutations demonstrated a similar performance. An outlier to this is the "RC: Smallest 1"-based permutations, which performed significantly poorer in comparison to the default Distance approach permutation. Apart from this case, deviation to accuracy presents a stable behaviour between both normalisation and cost-matrix utilisation when considered at first glance of these mean values. This could result in a misleading assumption being made, as the standard deviation of underlying FDI result values could demonstrate high variation between individual results by each permutation.

The isolation capability demonstrated by this range of functional parameter permutations demonstrates similar behaviour as observed for accuracy values. The default permutation for the Distance approach delivered an average rate of 36.67%, in correctly isolating operational conditions detected in the corresponding reference graphs. In addition to the case of the "RC: Smallest 1"-based permutations presenting significant deviation in their isolation values, three other permutations demonstrated notable increased performance in comparison to the default. These permutations include use of the "RC: Smallest 10" as cost-matrix utilisation, where isolation performance was increased by 6.6%, 5.7% and 10.9% in comparison to the default permutation. The MCC values in Table 5.5 indicate towards the NOC being consistently correctly detected, coupled with only moderate correct detection rates delivered by these Distance approach permutations. The standard deviation of individual samples for these performance measure values are further presented in Table 5.6, as supplemental information to illustrate the standard deviation of performance results delivered by each of the Distance approach permutations.

Table 5.5: Performance measures: Distance approach

	HEOM MEAN	HEOM RC: SMALLEST 1	HEOM RC: SMALLEST 5	HEOM RC: SMALLEST 10	HEOM Threshold: 2-SD MIN	HEOM Threshold: 3-SD MIN
ACCURACY (%)	76.667	39.524	75.714	78.571	77.143	76.190
ISOLATION (%)	36.667	6.190	30.952	43.333	35.238	35.238
MCC	0.377	0.164	0.366	0.384	0.394	0.371
	SHEOM MEAN	**SHEOM RC: SMALLEST 1**	**SHEOM RC: SMALLEST 5**	**SHEOM RC: SMALLEST 10**	**SHEOM Threshold: 2-SD MIN**	**SHEOM Threshold: 3-SD MIN**
ACCURACY (%)	78.571	40.000	76.190	78.571	82.381	77.619
ISOLATION (%)	36.190	6.667	30.952	42.381	37.619	37.143
MCC	0.394	0.166	0.371	0.385	0.433	0.379
	IQR-HEOM MEAN	**IQR-HEOM RC: SMALLEST 1**	**IQR-HEOM RC: SMALLEST 5**	**IQR-HEOM RC: SMALLEST 10**	**IQR-HEOM Threshold: 2-SD MIN**	**IQR-HEOM Threshold: 3-SD MIN**
ACCURACY (%)	84.762	37.619	71.905	78.095	77.143	82.381
ISOLATION (%)	38.095	7.143	31.905	47.619	38.095	36.667
MCC	0.455	0.108	0.338	0.381	0.367	0.424

Table 5.6: Standard deviation of performance measures: Distance approach

	HEOM MEAN	HEOM RC: SMALLEST 1	HEOM RC: SMALLEST 5	HEOM RC: SMALLEST 10	HEOM Threshold: 2-SD MIN	HEOM Threshold: 3-SD MIN
ACCURACY (%)	9.392	7.693	8.896	6.117	10.817	8.781
ISOLATION (%)	7.982	2.182	6.818	6.884	6.801	6.801
MCC	0.090	0.028	0.084	0.060	0.124	0.087
	SHEOM MEAN	**SHEOM RC: SMALLEST 1**	**SHEOM RC: SMALLEST 5**	**SHEOM RC: SMALLEST 10**	**SHEOM Threshold: 2-SD MIN**	**SHEOM Threshold: 3-SD MIN**
ACCURACY (%)	8.039	7.127	9.035	6.477	6.406	7.693
ISOLATION (%)	8.025	2.333	6.477	6.547	6.190	5.553
MCC	0.091	0.026	0.084	0.062	0.084	0.070
	IQR-HEOM MEAN	**IQR-HEOM RC: SMALLEST 1**	**IQR-HEOM RC: SMALLEST 5**	**IQR-HEOM RC: SMALLEST 10**	**IQR-HEOM Threshold: 2-SD MIN**	**IQR-HEOM Threshold: 3-SD MIN**
ACCURACY (%)	4.151	8.371	9.863	6.459	4.666	4.286
ISOLATION (%)	5.634	3.194	6.406	6.023	3.689	5.238
MCC	0.058	0.171	0.092	0.062	0.045	0.056

In summary, the cost-matrix utilisation as functional parameter presented both positive and negative deviation in comparison to the default functional permutation of the Distance approach. Deviation in performance between normalisation entries were observed to be minor in comparison to the deviation observed for the cost-matrix utilisation. It is noteworthy that the largest improvements to accuracy and isolation respectively were delivered by separate permutations instead of a single functional parameter permutation. While the IQR-HEOM MEAN delivered the highest accuracy, the highest isolation was demonstrated by the IQR-HEOM RC: SMALLEST 10. The standard deviation associated with each of these values were also smaller (better) than those observed for the default HEOM-MEAN permutation.

5.3.2 Sensitivity analysis: Distance approach

Following the execution of the above-mentioned FDI using the range of Distance approach permutations, individual FDI performance measure values were used as samples in performing the sensitivity analysis of the functional parameters selected. From the 10 datasets utilised, each section of the two-way ANOVA table contains 10 samples for each functional parameter combination, as seen in Table 5.7, populated with accuracy values for the respective Distance approach permutations.

Table 5.7: Accuracy values - Distance approach

ACCURACY		COST MATRIX UTILISATION					
		MEAN	RC: SMALLEST 1	RC: SMALLEST 5	RC: SMALLEST 10	Threshold: 2-SD MIN	Threshold: 3-SD MIN
NORMALISATION	HEOM	61.905	47.619	90.476	66.667	57.143	66.667
		76.190	42.857	85.714	80.952	76.190	76.190
		61.905	42.857	76.190	76.190	66.667	66.667
		85.714	33.333	66.667	85.714	85.714	85.714
		71.429	33.333	76.190	71.429	66.667	76.190
		90.476	52.381	76.190	76.190	85.714	76.190
		80.952	33.333	57.143	80.952	76.190	61.905
		71.429	47.619	76.190	76.190	76.190	76.190
		80.952	33.333	80.952	85.714	85.714	90.476
		85.714	28.571	71.429	85.714	95.238	85.714
	SHEOM	66.667	47.619	90.476	66.667	71.429	66.667
		76.190	42.857	85.714	80.952	80.952	76.190
		71.429	42.857	80.952	80.952	76.190	66.667
		90.476	33.333	66.667	85.714	90.476	85.714
		71.429	33.333	76.190	71.429	76.190	80.952
		90.476	52.381	76.190	71.429	85.714	85.714
		80.952	33.333	57.143	80.952	80.952	66.667
		71.429	47.619	76.190	76.190	80.952	80.952
		85.714	33.333	80.952	85.714	90.476	85.714
		80.952	33.333	71.429	85.714	90.476	80.952
	IQR-HEOM	85.714	47.619	71.429	66.667	71.429	80.952
		85.714	38.095	85.714	85.714	76.190	85.714
		80.952	47.619	66.667	71.429	71.429	80.952
		85.714	38.095	66.667	80.952	76.190	80.952
		76.190	42.857	61.905	76.190	76.190	76.190
		85.714	33.333	90.476	76.190	85.714	85.714
		90.476	28.571	66.667	71.429	80.952	85.714
		80.952	38.095	57.143	85.714	80.952	76.190
		85.714	19.048	76.190	80.952	80.952	80.952
		90.476	42.857	76.190	85.714	71.429	90.476

From the MCC values delivered by the range of Distance approach permutations, the ANOVA executed for this performance measure seen in Table 5.8 illustrates the layout of the table from which the various sums of squares values are calculated. These sums of squares calculated for the respective functional parameters, as the independent variables during analysis, have resulted in the following outcomes when comparing the respective F-values:

- The F-value for normalisation (Row) is less than the F-critical value, and indicates no significant difference in group means. The null hypothesis can be accepted for this source of variation.
- The F-value for cost-matrix utilisation (Column) is larger than the F-critical value, which is an indication of significant cause to variation being observed amongst groups. The null hypothesis for the cost-matrix utilisation is to be rejected, while the alternate hypothesis – that of significant difference in group means – is to be accepted.
- The F-value, when compared to the associated F-critical value, indicates to the null hypothesis being accepted to no significant difference being observed between group means.

Table 5.8: ANOVA MCC - Distance approach

Anova: Two-Factor With Replication — **Distance Approach - MCC**

SUMMARY	MEAN	RC: SMALLEST 1	RC: SMALLEST 5	RC: SMALLEST 10	THRESHOLD: 2-SD MIN	THRESHOLD: 2-SD MIN	Total
HEOM							
Sum	3.77	1.64	3.66	3.84	3.94	3.71	20.57
Average	0.38	0.16	0.37	0.38	0.39	0.37	0.34
Variance	0.01	8.89E-04	0.01	3.94E-03	0.02	0.01	0.01
SHEOM							
Sum	3.94	1.66	3.71	3.85	4.33	3.79	21.28
Average	0.39	0.17	0.37	0.39	0.43	0.38	0.35
Variance	0.01	7.73E-04	0.01	4.27E-03	0.01	0.01	0.01
IQR-HEOM							
Sum	4.55	1.08	3.38	3.81	3.67	4.24	20.73
Average	0.46	0.11	0.34	0.38	0.37	0.42	0.35
Variance	3.75E-03	0.03	0.01	4.34E-03	2.26E-03	3.52E-03	0.02
Columns							
Count	30	30	30	30	30	30	
Sum	12.27	4.38	10.74	11.51	11.94	11.74	
Average	0.41	0.15	0.36	0.38	0.40	0.39	
Variance	0.01	0.01	0.01	3.90E-03	0.01	0.01	
ANOVA							
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Row	4.66E-03	2	2.33E-03	0.30	0.74	3.05	TRUE
Column	1.51	5	0.30	39.27	2.73E-26	2.27	FALSE
Interaction	0.10	10	0.01	1.24	0.27	1.89	TRUE
Within	1.24	162	0.01				
Total	2.85	179					

Analysis of the accuracy values delivered by the range of Distance approach permutations presents a similar outcome as observed for the ANOVA of MCC values. The F-values calculated in analysing the accuracy values delivered the following outcomes:

- The F-value for normalisation (Row) is less than the F-critical value, and the null hypothesis can be accepted.
- The F-value for cost-matrix utilisation (Column) is larger than the F-critical value, and the alternate hypothesis – a significant difference exists in group means – should be accepted.
- The F-value, when compared to the associated F-critical value, indicates to the null hypothesis being accepted, with no significant difference being observed between group means.

Table 5.9: ANOVA Accuracy - Distance approach

Anova: Two-Factor With Replication — **Distance Approach - Accuracy**

SUMMARY	MEAN	RC: SMALLEST 1	RC: SMALLEST 5	RC: SMALLEST 10	THRESHOLD: 2-SD MIN	THRESHOLD: 2-SD MIN	Total
HEOM							
Sum	766.67	395.24	757.14	785.71	771.43	761.90	4238.10
Average	76.67	39.52	75.71	78.57	77.14	76.19	70.63
Variance	98.01	65.76	87.93	41.57	130.01	85.66	275.31
SHEOM							
Sum	785.71	400	761.90	785.71	823.81	776.19	4333.33
Average	78.57	40	76.19	78.57	82.38	77.62	72.22
Variance	71.81	56.44	90.70	46.61	45.60	65.76	272.24
IQR-HEOM							
Sum	847.62	376.19	719.05	780.95	771.43	823.81	4319.05
Average	84.76	37.62	71.90	78.10	77.14	82.38	71.98
Variance	19.15	77.85	108.09	46.36	24.19	20.41	302.16
Columns							
Count	30	30	30	30	30	30	
Sum	2400	1171.43	2238.10	2352.38	2366.67	2361.90	
Average	80	39.05	74.60	78.41	78.89	78.73	
Variance	71	63.18	92.79	41.81	68.31	60.57	
ANOVA							
Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***	***Null Hyphothesis***
Row	87.93	2	43.97	0.67	0.51	3.05	TRUE
Column	38688.46	5	7737.69	117.84	4.24E-52	2.27	FALSE
Interaction	807	10	80.70	1.23	0.28	1.89	TRUE
Within	10637.19	162	65.66				
Total	50220.58	179					

Table 5.10: ANOVA Isolation - Distance approach

Anova: Two-Factor With Replication — **Distance Approach - Isolation**

SUMMARY	MEAN	RC: SMALLEST 1	RC: SMALLEST 5	RC: SMALLEST 10	THRESHOLD: 2-SD MIN	THRESHOLD: 2-SD MIN	Total
HEOM							
Sum	366.67	61.90	309.52	433.33	352.38	352.38	1876.19
Average	36.67	6.19	30.95	43.33	35.24	35.24	31.27
Variance	70.80	5.29	51.65	52.66	51.40	51.40	184.76
SHEOM							
Sum	361.90	66.67	309.52	423.81	376.19	371.43	1909.52
Average	36.19	6.67	30.95	42.38	37.62	37.14	31.83
Variance	71.55	6.05	46.61	47.62	42.58	34.27	177.94
IQR-HEOM							
Sum	380.95	71.43	319.05	476.19	380.95	366.67	1995.24
Average	38.10	7.14	31.90	47.62	38.10	36.67	33.25
Variance	35.27	11.34	45.60	40.31	15.12	30.49	187.93
Columns							
Count	30	30	30	30	30	30	
Sum	1109.52	200	938.10	1333.33	1109.52	1090.48	
Average	36.98	6.67	31.27	44.44	36.98	36.35	
Variance	55.80	7.19	44.86	49.00	35.47	36.72	
ANOVA							
Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***	***Null Hyphothesis***
Row	125.72	2	62.86	1.59	0.21	3.05	TRUE
Column	25970.77	5	5194.15	131.68	3.48E-55	2.27	FALSE
Interaction	126.73	10	12.67	0.32	0.97	1.89	TRUE
Within	6390.02	162	39.44				
Total	32613.25	179					

The analysis of isolation values in Table 5.10 delivered by the Distance approach permutations present an identical case to what was previously observed for both MCC and accuracy. Only the cost-matrix utilisation (Column) demonstrated significant variation amongst group means, resulting in the acceptance of the alternate hypothesis for this source of variation.

The sensitivity demonstrated towards each of the respective sources of variation during analysis of each performance measure was consistent for the Distance approach. In calculating the sensitivity value towards each source of variation using (31), the sensitivity is demonstrated proportionately to the total sum of variation in Table 5.11 for MCC, accuracy, and isolation respectively. Figure 5.1 provides a visual demonstration to compare the scale of sensitivity towards each source of variation considered in the ANOVA performed. This sensitivity is calculated independently for each performance measure, and does not serve any influence towards sensitivity exhibited for the other performance measures.

From the sensitivity demonstrated in Table 5.11 below, sensitivity towards the normalisation functional parameter as source of variation is minimal for each of the performance measures considered. Thus, the Distance approach does present significant sensitivity towards normalisation as a functional parameter. Significantly higher sensitivity is demonstrated towards cost-matrix utilisation for each of the respective performance measures than any other source of variation in the ANOVA performed. This observation returns to the performance values seen in the previous section, where a clear difference was observed amongst column-wise cost-matrix utilisation selections. While minimal sensitivity was observed to the interaction between normalisation and cost-matrix utilisation, notable sensitivity was observed for the variation within ANOVA sample groups, as is indicative to minor performance variation in the FDI results for the datasets used in this study.

Table 5.11: Sensitivity values - Distance approach

Source of Variation	**MCC**	**Accuracy**	**Isolation**
Normalisation	0.002	0.002	0.004
Cost-Matrix Utilisation	0.529	0.770	0.796
Interaction	0.033	0.016	0.004
Within	0.436	0.212	0.196

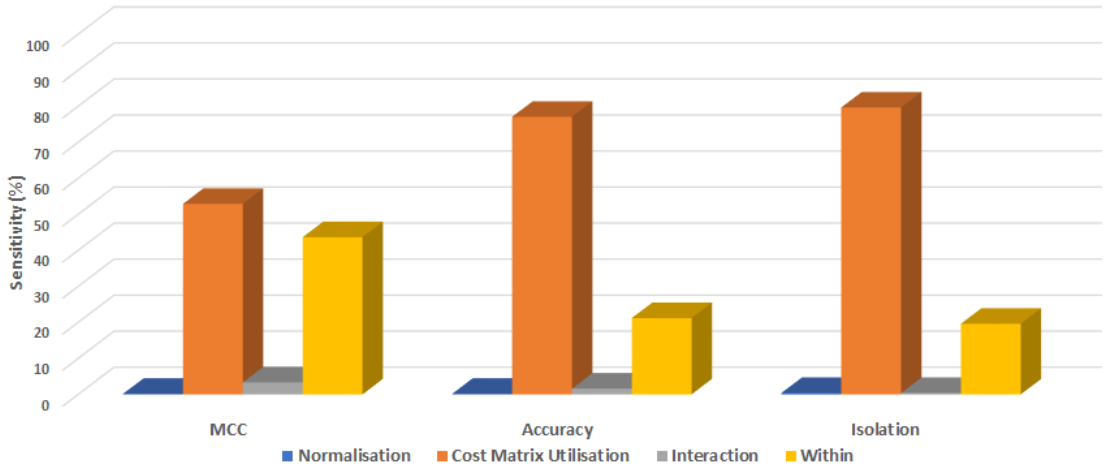

Figure 5.1: Performance measure sensitivity - Distance approach

5.4 Eigendecomposition - Qualitative approach

Execution of the Eigendecomposition – A Qualitative approach for FDI is presented in this section, as per the structure of functional parameter permutations demonstrated in Chapter 3, while utilising the same datasets as initially acquired for this study. This section will discuss the FDI results obtained for these functional parameter permutations according to the identified performance measures, and is accompanied by the ANOVA performed to determine whether significant sensitivity is demonstrated towards the functional parameters utilised.

5.4.1 FDI: Qualitative approach

Permutations of the Qualitative approach demonstrated variable capability in correctly detecting fault conditions observed, where Table 5.12 demonstrates results delivered by the default permutation HEOM 3-SD for dataset 1. While the NOC was correctly detected and isolated, only one fault condition was correctly isolated amongst those observed. Further inspection of the results indicates that the majority share of classifications were incorrectly assigned to the NOC instead of an alternate fault condition, which caused the low correct detection rate delivered for this dataset.

Table 5.12: FDI Results: HEOM 3-SD - Dataset 1

		REFERENCE GRAPH DATABASE																				FDI OUTCOME	
		NOC	FAULT 1	FAULT 2	FAULT 3	FAULT 4	FAULT 5	FAULT 6	FAULT 7	FAULT 8	FAULT 9	FAULT 10	FAULT 11	FAULT 12	FAULT 13	FAULT 14	FAULT 15	FAULT 16	FAULT 17	FAULT 18	FAULT 19	FAULT 20	DETECTION	ISOLATION
OPERATION GRAPHS	NOC	65.21	5.00	9.69	15.63	0	0	0	0	0.10	0.31	0.21	0	0	0	0	3.02	0.31	0.52	0	0	0	TrueNeg	Correct
	FAULT 1	19.69	77.29	0.31	0	0	0	0	0	0	0	0	0	0	0	2.08	0.63	0	0	0	0	0	TruePos	Correct
	FAULT 2	60.63	0	1.04	24.27	0	0	0	0	0	0.94	0	0	0	0.21	0.31	11.25	0.10	1.04	0	0	0.21	FalseNeg	Missed
	FAULT 3	47.81	0	2.19	46.46	0	0	0	0	0	1.15	0	0	0	0.42	0.42	0.63	0	0.73	0.21	0	0	FalseNeg	Missed
	FAULT 4	23.44	0.10	0	0	0	0	0	0	0	0	0.42	6.25	0.63	64.27	0.21	0	0	4.48	0.10	0	0.10	TruePos	Missed
	FAULT 5	74.27	0	0	2.08	0.10	21.98	0	0	0	0.21	0	0	0	1.15	0	0	0	0.21	0	0	0	FalseNeg	Missed
	FAULT 6	17.08	0	1.15	0	0	0	0	0	0.10	0	0.21	0	0	0	0	0	0	0	0	81.46	0	TruePos	Missed
	FAULT 7	20.10	0.10	0	0	0	0	0	0	33.96	0	26.04	0	15.94	0.31	0.10	0.52	0	0	0	1.77	1.15	TruePos	Missed
	FAULT 8	67.60	0	3.13	16.25	0	0	0	0	0	8.33	1.25	0	0	0	0.63	1.88	0.94	0	0	0	0	FalseNeg	Missed
	FAULT 9	79.90	0	7.71	0	0	0	0	0	0	0.31	0	0	0	0.21	0.10	11.25	0.21	0.31	0	0	0	FalseNeg	Missed
	FAULT 10	19.17	0	0	0	0	0	0	0.10	0	0	32.81	0	0.21	0.10	0	0	46.98	0.21	0.21	0.10	0.10	TruePos	Missed
	FAULT 11	16.77	0	0	0	83.23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	TruePos	Missed
	FAULT 12	16.88	0	80.73	0.31	0.10	0	0	0	0	1.25	0	0	0	0.10	0	0	0	0.31	0.31	0	0	TruePos	Missed
	FAULT 13	17.19	0	0	0	45.31	0	0	0	0	0	0	0.21	0	0	0	0	0	37.29	0	0	0	TruePos	Missed
	FAULT 14	76.25	0.10	0	0	0	0	0	0	2.50	9.90	0	0	0	2.81	0	6.98	0	0	0.10	1.35	0	FalseNeg	Missed
	FAULT 15	28.85	0.10	0	34.38	0	0	0	0	0	6.25	0	0.10	0	0.10	0.63	7.71	0.10	0.94	0.10	20.31	0.42	TruePos	Missed
	FAULT 16	49.17	0	0	0	0	0	0	0	0	0	10.00	0	0	0	0	0	40.83	0	0	0	0	FalseNeg	Missed
	FAULT 17	36.98	0.10	0	10.10	0	0	0	0	10.31	12.92	0	0.31	0.73	0.21	0	1.25	0	26.25	0	0.10	0.73	FalseNeg	Missed
	FAULT 18	16.67	0.21	0.21	6.25	0	0.52	0	0	0	0.31	0	25.83	0	49.48	0.21	0	0	0.31	0	0	0	TruePos	Missed
	FAULT 19	65.21	0	0	22.81	0	0	0	0	0	0.52	0.21	0	0	0.10	0.21	10.63	0.10	0.21	0	0	0	FalseNeg	Missed
	FAULT 20	63.54	0	0	3.54	0	0	0	0	0	0	0	0.73	0	0	0	0	0	0.83	0	0	31.35	FalseNeg	Missed

From results illustrated in Tables 5.13 and 5.14, the SHEOM 1-SD and IQR-HEOM 2-MAD respectively demonstrate a similar to higher capability in delivering correct detection classifications, while instead delivering a notable improvement to isolation classifications per the datasets observed. A similar case is observed in the majority of incorrect classifications being allocated to the NOC, with the IQR-HEOM 2-MAD presenting a minor improvement towards discerning between normal and fault condition when inspecting values for faults 11-13, and faults 16-20 in Table 5.14 below.

Table 5.13: FDI Results: SHEOM 1-SD - Dataset 5

		REFERENCE GRAPH DATABASE																				FDI OUTCOME	
		NOC	FAULT 1	FAULT 2	FAULT 3	FAULT 4	FAULT 5	FAULT 6	FAULT 7	FAULT 8	FAULT 9	FAULT 10	FAULT 11	FAULT 12	FAULT 13	FAULT 14	FAULT 15	FAULT 16	FAULT 17	FAULT 18	FAULT 19	FAULT 20	DETECTION	ISOLATION
OPERATION GRAPHS	NOC	51.35	0	43.85	0	0	0	0	0	0	1.46	0.31	0	0	0.10	0.10	1.77	0.21	0.73	0.10	0	0	TrueNeg	Correct
	FAULT 1	50.31	49.17	0.10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.10	0.10	0.10	FalseNeg	Missed
	FAULT 2	27.92	0	62.60	0	0	0	0	0	0	2.40	0.31	0	0	0.10	1.35	5.00	0.10	0.21	0	0	0	TruePos	Correct
	FAULT 3	20.21	0	0	69.27	0	0	0	0	0	6.15	0.63	0	0	0	0	2.92	0.10	0.10	0	0.63	0	TruePos	Correct
	FAULT 4	19.79	1.15	0	0	18.96	0	0	0	0	0.21	0	59.90	0	0	0	0	0	0	0	0	0	TruePos	Missed
	FAULT 5	17.19	0	0	0	0	82.50	0	0	0.10	0.10	0	0.10	0	0	0	0	0	0	0	0	0	TruePos	Correct
	FAULT 6	21.77	0.21	0	0	2.81	0	0.10	0	60.73	1.98	8.54	0	1.98	0.73	0.21	0.73	0.10	0	0	0	0.10	TruePos	Missed
	FAULT 7	16.67	0	80.10	0	0.73	0	0.10	0.10	0.94	0	0.42	0	0	0	0.52	0	0.10	0	0	0	0.31	TruePos	Missed
	FAULT 8	60.42	0	0	14.69	0	0	0	0	0.10	7.92	0	0.83	3.02	7.81	1.25	2.40	0	1.15	0	0	0.42	FalseNeg	Missed
	FAULT 9	64.90	0	0	0.52	0	0	0	0	0	3.85	0.63	0	0	0.42	0.21	2.81	0.21	0.52	0.21	25.73	0	FalseNeg	Missed
	FAULT 10	60.31	0.10	36.04	0	0	0	0	0	0.10	0.31	0.52	0	0.73	0	0	0.73	0	0	0	1.15	0	FalseNeg	Missed
	FAULT 11	28.75	64.58	0	0	0	0	0	0.10	0.10	0	0	6.04	0.42	0	0	0	0	0	0	0	0	TruePos	Missed
	FAULT 12	25.42	0	0	0	0	0	0	0	0.10	1.25	0.42	0	0.21	0.10	0	7.19	18.85	0.21	0	46.25	0	TruePos	Missed
	FAULT 13	85.42	0	3.23	0	0	0	0	0	0.31	1.88	0.63	0	0	0.10	0.94	7.08	0	0.10	0.31	0	0	FalseNeg	Missed
	FAULT 14	50.42	0	0.10	0	0	0	0	0	0	0.52	38.02	0	0	0.21	0.42	0	1.46	0	8.65	0.21	0	FalseNeg	Missed
	FAULT 15	52.60	0	44.06	0	0	0	0	0	0.10	1.15	0.42	0.10	0	0.21	0.10	0.21	0.10	0.52	0.10	0.10	0.21	FalseNeg	Missed
	FAULT 16	99.58	0	0	0	0	0	0	0	0	0	0.42	0	0	0	0	0	0	0	0	0	0	FalseNeg	Missed
	FAULT 17	95.73	0	0	0.21	0	0	0	0	0	0.10	0	0.10	0.94	0	0	1.56	0.42	0	0	0	0.94	FalseNeg	Missed
	FAULT 18	56.15	0	37.50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6.35	FalseNeg	Missed
	FAULT 19	17.29	0.31	1.04	0	0	0	0	0	0.10	1.25	0.21	0.10	0.10	0.21	0.10	1.98	0.10	0.10	0	77.08	0	TruePos	Correct
	FAULT 20	52.81	0	6.15	0.42	0	0	0	0	0.52	6.25	0.31	11.04	0.31	0.52	0.10	0	20.31	0	0	0	1.25	FalseNeg	Missed

Table 5.14: FDI Results: IQR-HEOM 2-MAD - Dataset 8

		REFERENCE GRAPH DATABASE																					FDI OUTCOME	
		NOC	FAULT 1	FAULT 2	FAULT 3	FAULT 4	FAULT 5	FAULT 6	FAULT 7	FAULT 8	FAULT 9	FAULT 10	FAULT 11	FAULT 12	FAULT 13	FAULT 14	FAULT 15	FAULT 16	FAULT 17	FAULT 18	FAULT 19	FAULT 20	DETECTION	ISOLATION
OPERATION GRAPHS	NOC	92.71	0	0	0.73	0	0.10	0	0	0.21	0	0	0	0	0.31	0	4.17	0	0.31	0.31	0.42	0.73	TrueNeg	Correct
	FAULT 1	17.60	82.19	0.10	0	0	0	0	0	0	0	0	0	0	0	0	0.10	0	0	0	0	0	TruePos	Correct
	FAULT 2	32.50	0	65.42	0	0	0	0	0	0	0.21	0.10	0	0	0.42	0	0.10	0.10	0.31	0.31	0	0.52	TruePos	Correct
	FAULT 3	18.23	0	0	80.21	0	0	0	0	0	1.56	0	0	0	0	0	0	0	0	0	0	0	TruePos	Correct
	FAULT 4	16.67	0	0	9.48	0	0.10	0	0	0	0	0	31.56	0	0	0	0	0	42.19	0	0	0	TruePos	Missed
	FAULT 5	86.56	0	12.60	0	0	0	0	0	0	0	0	0	0	0.10	0.10	0.42	0	0	0.21	0	0	FalseNeg	Missed
	FAULT 6	26.67	13.54	0	9.90	0	0	0	0.10	3.44	29.17	0	0.21	0.21	0	0	0.21	0	0	10.94	4.90	0.73	TruePos	Missed
	FAULT 7	90.94	0	0	0	0	0	0	0	3.65	0	0	0	0.10	0.21	0	5.10	0	0	0	0	0	FalseNeg	Missed
	FAULT 8	86.46	0	0	0	0	0	0	0	13.54	0	0	0	0	0	0	0	0	0	0	0	0	FalseNeg	Missed
	FAULT 9	20.83	0	0.63	76.56	0	0	0	0	0	1.77	0	0	0	0.21	0	0	0	0	0	0	0	TruePos	Missed
	FAULT 10	92.81	0	0	0	0	0.10	0	0	0.31	0.10	0.10	0	0.10	0.10	0.63	2.50	0.10	0.52	1.77	0.21	0.63	FalseNeg	Missed
	FAULT 11	17.40	20.73	0	0.10	0	0	0.21	0	0	0	0	22.81	14.17	0.10	0	0.63	17.29	0	0.10	0	6.46	TruePos	Correct
	FAULT 12	16.67	0.94	0.10	0	0	0.10	0	0	26.04	0	0	0	31.04	0	25.00	0	0.10	0	0	0	0	TruePos	Correct
	FAULT 13	26.25	0	0	2.19	0.10	0	0	0	0.42	6.98	0	0.10	0	0.31	0	2.92	0.10	0.42	0	60.00	0.21	TruePos	Missed
	FAULT 14	86.25	0	0.10	0	0	0	0	0	0.21	0.73	0.10	0.10	0	0.31	0.31	3.54	0	0	0.21	7.50	0.63	FalseNeg	Missed
	FAULT 15	97.19	0	0	0	0.10	0	0	0	0	0.52	0	0	0	0	0.10	1.77	0	0.21	0	0.10	0	FalseNeg	Missed
	FAULT 16	16.77	0.10	0.21	0	0	50.63	0.10	0	0	0.73	0	0.10	0.10	0	0	0	31.25	0	0	0	0	TruePos	Missed
	FAULT 17	21.77	0	0.21	0	0	0	0	0	0	16.98	0	8.96	0	0.52	0	0	0.10	51.35	0	0	0.10	TruePos	Correct
	FAULT 18	48.85	0	15.31	17.29	0	0	0	0	0	16.46	0	0	0	1.04	0	0.52	0	0	0.52	0	0	FalseNeg	Missed
	FAULT 19	41.46	0	55.52	0	0	0	0	0	0.21	0.21	0	0.10	0	0.42	0	1.04	0	0.52	0.10	0.21	0.21	TruePos	Missed
	FAULT 20	16.77	0	0.21	0	0.10	1.46	0.10	0	0	0	20.42	5.94	0.31	21.88	0	0	0	0	0	0	32.81	TruePos	Correct

Performance results delivered by the respective permutations of the Qualitative approach present a significant change in conditions in comparison to results observed for the Distance approach above. The default HEOM 3-SD permutation demonstrated only moderate accuracy to correctly classify conditions observed, and a considerably lower isolation rate as extension of the accuracy observed. The default permutation presented both the lowest accuracy and isolation rate amongst the range of Qualitative approach permutations, with only a single case presenting 1% difference for each of these measures. The HEOM 2-MAD demonstrated similar accuracy to the default permutation, while demonstrating a 5% improvement to isolation over the default permutation. The IQR-HEOM 3-SD presents a contrast, where a 13.8% improvement to accuracy is delivered in comparison to the default permutation, while only yielding 0.5% difference from the default permutation's isolation rate.

The utilisation of the MCC as performance measure now presents a clear case to the importance in monitoring an FDI method's capability to correctly classify the normal condition when evaluating performance. MCC values observed in Table 5.15 for the Qualitative approach permutations illustrate four cases where moderate accuracy is coupled with a negative MCC value delivered by a permutation. These negative MCC values are indicative of the respective permutations failing to correctly discern the normal condition from the range of fault conditions observed, while still presenting moderate accuracy results. The highest accuracy value was delivered by the IQR-HEOM 2-SD permutation, while the highest isolation value was delivered by the IQR-HEOM 1-SD, with an increase of 11% when compared to the default HEOM 3-SD.

Table 5.16 presents the standard deviation values associated with the performance results delivered by the respective Qualitative approach permutations. This provides additional context when comparing deviation in performance between the respective permutations and the default HEOM-3SD. This can be observed for the HEOM 1-MAD, where improvement to performance delivered is associated with a larger standard deviation of the overall results delivered by this permutation.

Table 5.15: Performance measures - Qualitative approach

	HEOM 3-SD	HEOM 2-SD	HEOM 1-SD	HEOM 2-MAD	HEOM 1-MAD
ACCURACY (%)	50.952	57.143	58.571	51.905	55.238
ISOLATION (%)	14.286	19.524	20.476	19.524	20.476
MCC	0.210	0.237	0.246	0.215	0.236
	SHEOM 3-SD	**SHEOM 2-SD**	**SHEOM 1-SD**	**SHEOM 2-MAD**	**SHEOM 1-MAD**
ACCURACY (%)	62.381	58.571	60.476	57.143	63.810
ISOLATION (%)	19.524	24.286	20.476	24.286	23.810
MCC	-0.158	-0.002	0.173	0.158	0.232
	IQR-HEOM 3-SD	**IQR-HEOM 2-SD**	**IQR-HEOM 1-SD**	**IQR-HEOM 2-MAD**	**IQR-HEOM 1-MAD**
ACCURACY (%)	64.762	65.238	55.714	56.667	55.238
ISOLATION (%)	14.762	18.571	25.714	24.286	17.143
MCC	-0.149	0.154	0.232	0.238	0.230

Table 5.16: Standard deviation of performance measures - Qualitative approach

	HEOM 3-SD	HEOM 2-SD	HEOM 1-SD	HEOM 2-MAD	HEOM 1-MAD
ACCURACY (%)	9.536	6.389	8.262	10.313	13.669
ISOLATION (%)	6.734	5.408	5.654	5.813	6.042
MCC	0.039	0.030	0.045	0.044	0.071
	SHEOM 3-SD	**SHEOM 2-SD**	**SHEOM 1-SD**	**SHEOM 2-MAD**	**SHEOM 1-MAD**
ACCURACY (%)	9.147	7.392	9.048	8.518	6.801
ISOLATION (%)	4.972	3.956	6.406	6.547	6.023
MCC	0.033	0.226	0.177	0.181	0.141
	IQR-HEOM 3-SD	**IQR-HEOM 2-SD**	**IQR-HEOM 1-SD**	**IQR-HEOM 2-MAD**	**IQR-HEOM 1-MAD**
ACCURACY (%)	8.571	6.406	8.262	10.090	9.571
ISOLATION (%)	6.190	5.408	6.098	6.884	4.364
MCC	0.031	0.204	0.041	0.049	0.044

5.4.2 Sensitivity analysis: Qualitative approach

Analysis of the MCC values delivered by Qualitative approach permutations presented in Table 5.17 illustrates an interesting case to where multiple sources were identified as significant cause to variation amongst group means. Both normalisation and cost-matrix utilisation presented significant variation between their respective sample groups, while additionally presenting significant interaction between these independent variables. This interaction is a result of clear difference in performance values, as these are more widely distributed per their respective ranges as opposed to the consistent results observed for the Distance approach. Note the size of the F-values calculated in comparison to their respective F-critical values, as this gives an indication of the significance to variation observed. Additional observation is made to the sums of squares to variation within samples observed. This constitutes a significant share of the total sums of squares indicating widespread variation amongst samples per respective sample group, and results in inconsistent or widely distributed results delivered by the respective Qualitative approach permutations.

Table 5.17: ANOVA MCC - Qualitative approach

Anova: Two-Factor With Replication — **Qualitative Approach - MCC**

SUMMARY	3-SD	2-SD	1-SD	2-MAD	1-MAD	Total
HEOM						
Sum	2.10	2.37	2.46	2.15	2.36	11.44
Average	0.21	0.24	0.25	0.22	0.24	0.23
Variance	1.70E-03	1.00E-03	2.22E-03	2.14E-03	5.62E-03	2.52E-03
SHEOM						
Sum	-1.58	-0.02	1.73	1.58	2.32	4.03
Average	-0.16	-1.85E-03	0.17	0.16	0.23	0.08
Variance	1.23E-03	0.06	0.03	0.04	0.02	0.05
IQR-HEOM						
Sum	-1.49	1.54	2.32	2.38	2.30	7.04
Average	-0.15	0.15	0.23	0.24	0.23	0.14
Variance	1.04E-03	0.05	1.85E-03	2.64E-03	2.18E-03	0.03
Columns						
Count	30	30	30	30	30	
Sum	-0.96	3.88	6.50	6.10	6.98	
Average	-0.03	0.13	0.22	0.20	0.23	
Variance	0.03	0.04	0.01	0.01	0.01	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Row	0.55	2	0.28	19.11	4.94E-08	3.06	FALSE
Column	1.43	4	0.36	24.62	2.59E-15	2.44	FALSE
Interaction	0.68	8	0.09	5.87	1.94E-06	2.01	FALSE
Within	1.96	135	0.01				
Total	4.63	149					

Analysis of the accuracy results in Table 5.18 delivered by the range of method permutations indicates only the normalisation (Row) presenting significant variation to sample group means. The null hypothesis associated with this functional parameter as a source of variation not being significant should be rejected, where the alternate hypothesis should be accepted as this functional parameter is a significant cause to variation observed for the group means. The within group additionally presented the majority share to the total sums of squares observed, and is again indicative of samples being widely distributed within the respective sample groups as a result of inconsistent behaviour by the respective permutations.

The isolation results ANOVA delivered by the range of permutations in Table 5.19 indicates both normalisation and cost-matrix utilisation, presenting significant variation to group means respectively. The null hypothesis for both parameters can be rejected and their alternate hypotheses are to be accepted based on their F-values exceeding the respective F-critical values. No significant variation was observed for the interaction between independent variables, while a significant variation within sample groups maintains its presence as observed for the previous performance measures.

Table 5.18: ANOVA Accuracy - Qualitative approach

Anova: Two-Factor With Replication — **Qualitative Approach - Accuracy**

SUMMARY	3-SD	2-SD	1-SD	2-MAD	1-MAD	Total
HEOM						
Sum	509.52	571.43	585.71	519.05	552.38	2738.10
Average	50.95	57.14	58.57	51.90	55.24	54.76
Variance	101.03	45.35	75.84	118.17	207.61	109.45
SHEOM						
Sum	623.81	585.71	604.76	571.43	638.10	3023.81
Average	62.38	58.57	60.48	57.14	63.81	60.48
Variance	92.97	60.72	90.95	80.62	51.40	75.20
IQR-HEOM						
Sum	647.62	652.38	557.14	566.67	552.38	2976.19
Average	64.76	65.24	55.71	56.67	55.24	59.52
Variance	81.63	45.60	75.84	113.13	101.79	97.41
Columns						
Count	30	30	30	30	30	
Sum	1780.95	1809.52	1747.62	1657.14	1742.86	
Average	59.37	60.32	58.25	55.24	58.10	
Variance	123.13	59.95	79.26	102.59	128.86	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Row	937.26	2	468.63	5.24	0.01	3.06	FALSE
Column	438.40	4	109.60	1.22	0.30	2.44	TRUE
Interaction	1298.56	8	162.32	1.81	0.08	2.01	TRUE
Within	12083.90	135	89.51				
Total	14758.13	149					

Table 5.19: ANOVA Isolation - Qualitative approach

Anova: Two-Factor With Replication — **Qualitative Approach - Isolation**

SUMMARY	3-SD	2-SD	1-SD	2-MAD	1-MAD	Total
HEOM						
Sum	142.86	195.24	204.76	195.24	204.76	942.86
Average	14.29	19.52	20.48	19.52	20.48	18.86
Variance	50.39	32.50	35.53	37.54	40.56	41.61
SHEOM						
Sum	195.24	242.86	204.76	242.86	238.10	1123.81
Average	19.52	24.29	20.48	24.29	23.81	22.48
Variance	27.46	17.38	45.60	47.62	40.31	37.06
IQR-HEOM						
Sum	147.62	185.71	257.14	242.86	171.43	1004.76
Average	14.76	18.57	25.71	24.29	17.14	20.10
Variance	42.58	32.50	41.32	52.66	21.16	53.02
Columns						
Count	30	30	30	30	30	
Sum	485.71	623.81	666.67	680.95	614.29	
Average	16.19	20.79	22.22	22.70	20.48	
Variance	43.16	32.03	44.31	47.98	39.33	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Row	338.32	2	169.16	4.49	0.01	3.06	FALSE
Column	793.65	4	198.41	5.27	5.68E-04	2.44	FALSE
Interaction	573.24	8	71.66	1.90	0.06	2.01	TRUE
Within	5086.17	135	37.68				
Total	6791.38	149					

A summary of the sensitivity calculated for the respective performance measures is presented in Table 5.20 for each of the respective sources of variation considered in the ANOVA. Figure 5.2 presents a visual comparison to these sensitivity values, where a clear consistency is demonstrated in the sensitivity to variation within samples. This reinforces the observation that the performance by the Qualitative approach permutation is inconsistent and varies in its ability to classify conditions observed. Due to the size of the within sum of squares calculated for each of the ANOVAs performed, sensitivity to the within variation then forms the majority share of overall sensitivity observed in Figure 5.2, while individual significant sources to variation determined in the ANOVA are not clear in this figure. Thus, individual sensitivity should be solely based on the ANOVA hypotheses accepted, as the variation within sample groups delivered by the Qualitative approach's permutations are significantly distributed, resulting in inconclusive results if only based on the sensitivity values in Table 5.20.

Table 5.20: Sensitivity values - Qualitative approach

Source of Variation	MCC	Accuracy	Isolation
Normalisation	0.120	0.064	0.050
Cost-Matrix Utilisation	0.309	0.030	0.117
Interaction	0.147	0.088	0.084
Within	0.424	0.819	0.749

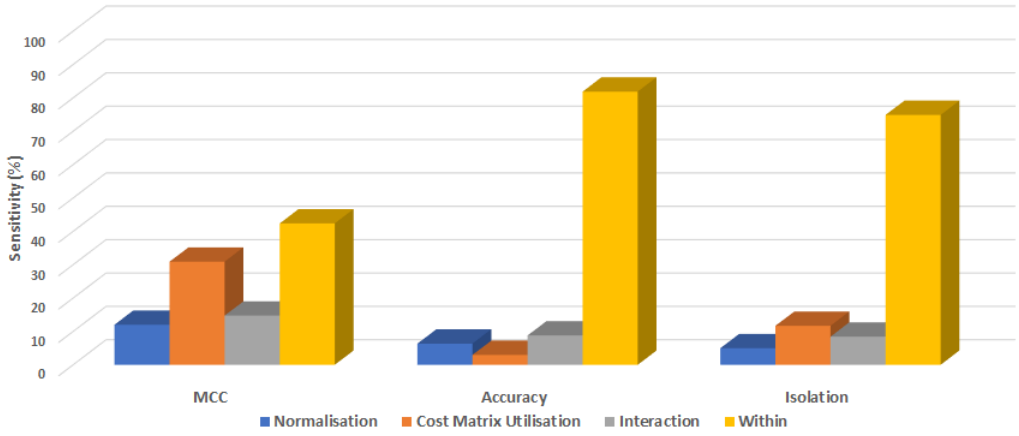

Figure 5.2: Performance measure sensitivity - Qualitative approach

5.5 Eigendecomposition - Quantitative approach

The execution of the Eigendecomposition - Quantitative approach for FDI – is presented in this section as per the three functional parameter permutations identified in Chapter 3 in utilising the datasets acquired for this study. This section will discuss the FDI results obtained for the respective permutations according to the identified performance measures, and is accompanied by the ANOVA performed to determine whether significant sensitivity is demonstrated towards normalisation as a functional parameter for this FDI method.

5.5.1 FDI: Quantitative approach

Results for the default HEOM permutation is presented in Table 5.21, where a large quantity of conditions were correctly detected, including the NOC, where only few conditions were correctly isolated. The detection and isolation delivered by the SHEOM permutation seen in Table 5.22 demonstrate a fair detection capability, with only a minor quantity of correct isolation classifications for the range of conditions observed.

Table 5.21: FDI Results: HEOM MEAN - Dataset 2

		REFERENCE GRAPH DATABASE																					FDI OUTCOME	
		NOC	FAULT 1	FAULT 2	FAULT 3	FAULT 4	FAULT 5	FAULT 6	FAULT 7	FAULT 8	FAULT 9	FAULT 10	FAULT 11	FAULT 12	FAULT 13	FAULT 14	FAULT 15	FAULT 16	FAULT 17	FAULT 18	FAULT 19	FAULT 20	DETECTION	ISOLATION
OPERATION GRAPHS	NOC	94.27	0	0.83	0	0	0	0	0	0.10	0	0.31	0.21	0	0.31	0.21	0	0.10	0.10	3.02	0	0.52	TrueNeg	Correct
	FAULT 1	16.67	0.42	4.90	0	0	0	0	0	0.10	0	0	0	0	77.92	0	0	0	0	0	0	0	TruePos	Missed
	FAULT 2	25.31	0	26.25	0	0	0	0	0	0.10	0	0.10	0.10	0	46.25	0	0.21	0.31	0.10	0.63	0	0.63	TruePos	Missed
	FAULT 3	19.06	0	0	75.73	0	0	0	0	0	0.73	0	0	0	0.10	0	2.29	0.63	0	1.35	0	0.10	TruePos	Correct
	FAULT 4	17.60	0	0.10	0	0.10	0	0	0	5.63	6.35	0	44.38	5.10	11.56	0	5.31	0	0	0	0.21	3.65	TruePos	Missed
	FAULT 5	71.15	0	0	0	0	0.10	0	0	0	0.63	0	0	19.48	0.52	0	4.58	0.21	0.31	3.02	0	0	FalseNeg	Missed
	FAULT 6	16.67	77.81	0.21	0	0	0	0	0	0	5.21	0.10	0	0	0	0	0	0	0	0	0	0	TruePos	Missed
	FAULT 7	16.67	0	0.63	0	0	0	0	0	0.21	0	0	0	0	0	81.98	0	0	0	0	0.52	0	TruePos	Missed
	FAULT 8	16.67	0	27.50	0	0	0	0	0	0.21	0	0.10	0	0	55.42	0	0	0	0	0	0	0.10	TruePos	Missed
	FAULT 9	18.13	0	76.35	0	0	0	0	0	0	1.15	0.31	0	0	1.15	0.10	1.35	0.10	0	0.21	0.94	0.21	TruePos	Missed
	FAULT 10	16.67	0	0	0	0	0	0	0	0.10	6.15	0.31	0	0	0	0.73	64.17	0.10	0	0	11.77	0	TruePos	Missed
	FAULT 11	16.67	0	0	0	0.10	0	0	0	0.10	0.10	0.10	24.58	0	0	57.50	0	0	0	0	0	0.83	TruePos	Missed
	FAULT 12	90.63	0	0	0	0	0	0	0	0.21	0.94	0.21	0	1.56	0.94	0.10	1.77	0.10	0.21	2.71	0	0.63	FalseNeg	Missed
	FAULT 13	16.67	0	0.21	0	0	0	0	0	0	1.67	0.31	0	0	0	44.38	4.06	0	0	0	32.71	0	TruePos	Missed
	FAULT 14	24.48	0	0	0	0	0	0	0	0.21	0	0.21	0	0	0	70.00	3.96	0.10	0	0.10	0.63	0.31	TruePos	Correct
	FAULT 15	61.15	0	0	0	0	0	0	0	0	0.10	0	0	0	29.38	0	5.83	0.94	0.10	2.08	0	0.42	FalseNeg	Missed
	FAULT 16	32.92	0	64.48	0	0	0	0	0	0	0	0.31	0.10	0	0.10	0.10	0	0.21	0.10	1.15	0	0.52	TruePos	Missed
	FAULT 17	33.65	0	0	0	0	0	0	0	0.10	0.42	0.10	0.83	0	0	60.83	0.42	0	0	0.73	0	2.92	TruePos	Missed
	FAULT 18	16.67	0	0	0	0	0	0	0	0.42	0.10	0.31	0	0.10	0	5.94	0.73	0	0.10	75.42	0.21	0	TruePos	Correct
	FAULT 19	18.44	0	0	0	0	0	0	0	0.10	0.21	0.42	0	0	0	26.25	3.96	0	0	0.21	50.42	0	TruePos	Correct
	FAULT 20	57.50	0	0	0	0	0	0	0	0	0.94	0	0	0	0.63	0	3.85	1.04	0.10	2.08	0.31	33.54	FalseNeg	Missed

The IQR-HEOM MEAN permutation demonstrated in Table 5.23 presents an interesting case to FDI results observed. While the initial observation of the detection classifications indicates a moderately high rate of correct detection classifications, caution should be taken, as the incorrect detection classification of the normal condition should be noted carefully. This "False Positive" classification is indicative of this permutation being unable to realistically discern between normal and possible fault conditions observed. The IQR-HEOM MEAN additionally presents a case of multiple isolation classifications, in order to select fault conditions indicative of this permutation perceiving these conditions as highly similar to the respective reference conditions.

Table 5.22: FDI Results: SHEOM MEAN - Dataset 5

		REFERENCE GRAPH DATABASE																				FDI OUTCOME	
		NOC	FAULT 1	FAULT 2	FAULT 3	FAULT 4	FAULT 5	FAULT 6	FAULT 7	FAULT 8	FAULT 9	FAULT 10	FAULT 11	FAULT 12	FAULT 13	FAULT 14	FAULT 15	FAULT 16	FAULT 17	FAULT 18	FAULT 19	FAULT 20	DETECTION	ISOLATION
OPERATION GRAPHS	NOC	49.48	0	47.50	0	0	0	0	0	0.21	0.31	0.10	0	0	0.21	0	0.63	0.10	1.15	0.21	0.10	0	TrueNeg	Correct
	FAULT 1	16.67	82.40	0.10	0	0	0	0	0	0.63	0	0	0	0.10	0	0.10	0	0	0	0	0	0	TruePos	Correct
	FAULT 2	19.69	0	79.48	0	0	0	0	0	0.10	0.10	0	0	0	0	0.10	0.42	0	0.10	0	0	0	TruePos	Correct
	FAULT 3	21.04	0	0	41.67	0	0	0	0	0	3.44	0	0	0	0.10	0	0.31	0.10	0.21	0.52	32.50	0.10	TruePos	Correct
	FAULT 4	16.67	0	0	0	83.33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	TruePos	Correct
	FAULT 5	16.77	0	0	0	0	82.81	0	0	0	0	0	0	0.10	0	0.21	0	0	0.10	0	0	0	TruePos	Correct
	FAULT 6	16.67	14.27	1.25	0	0	0	0.10	0	37.40	0.10	0	0	0.21	1.77	24.69	0	0.21	0	0.10	0	3.23	TruePos	Missed
	FAULT 7	16.67	0	2.71	0	0	0	0	0	0	0	0	0	0.10	0	78.65	0	0.52	0	0.31	0	1.04	TruePos	Missed
	FAULT 8	16.67	0	79.27	0	0	0	0	0	0	0.21	0	0	0	3.13	0	0.42	0	0	0	0.31	0	TruePos	Missed
	FAULT 9	58.13	0	0	0	0	0	0	0	0.10	1.77	0.10	0.10	0	0.42	0	1.35	0.21	0.94	0.63	36.25	0	FalseNeg	Missed
	FAULT 10	16.67	0	0	0	0	0	0	0	0.52	0.73	0	0	1.04	0.21	69.79	9.38	1.04	0	0.10	0.31	0.21	TruePos	Missed
	FAULT 11	22.40	0	0	0	0	0	0	0	0.31	0.21	0	0	0	0.31	0.31	0.21	0	0	0	75.73	0.52	TruePos	Missed
	FAULT 12	51.15	0	0	0	0	0	0	0	0.21	1.04	0	0	3.23	0.21	0.42	6.88	0.21	0	0.21	35.94	0.52	FalseNeg	Missed
	FAULT 13	73.13	0	21.46	0	0	0	0	0	0.10	0.10	0	0.10	0	0	0.10	3.23	0.10	1.56	0.10	0	0	FalseNeg	Missed
	FAULT 14	81.35	0	13.33	0	0	0	0	0	0.21	0	0	0	0	0.21	1.15	1.46	0.21	0.10	1.04	0.10	0.83	FalseNeg	Missed
	FAULT 15	52.50	0	45.31	0	0	0	0	0	0	0.10	0.10	0	0	0.42	0	0.42	0.10	0.73	0.10	0.10	0.10	FalseNeg	Missed
	FAULT 16	89.90	0	7.60	0	0	0	0	0	0.10	0.31	0.10	0	0	0.21	0.31	0	0.42	0.73	0.10	0	0.21	FalseNeg	Missed
	FAULT 17	16.67	0	0	0	0	0	0	0	0.42	0	0.94	0	8.13	0.31	9.58	43.75	0.83	0	0.21	0	19.17	TruePos	Missed
	FAULT 18	71.56	0	0	0	0	0	0	0	0.21	0.10	0.10	0	0	0.31	0.10	1.25	0.63	0.10	1.35	23.33	0.94	FalseNeg	Missed
	FAULT 19	18.54	0.21	72.29	0.10	0	0	0	0	0.10	0.21	0	0	0	0.31	0	0.83	0.10	0	0	7.19	0.10	TruePos	Missed
	FAULT 20	16.67	0	0	0	0	0	0	0	0.31	0	0	0	0	0.21	0.21	0.10	0	0	0.21	79.69	2.60	TruePos	Missed

Table 5.23: FDI Results: IQR-HEOM MEAN - Dataset 7

		REFERENCE GRAPH DATABASE																				FDI OUTCOME	
		NOC	FAULT 1	FAULT 2	FAULT 3	FAULT 4	FAULT 5	FAULT 6	FAULT 7	FAULT 8	FAULT 9	FAULT 10	FAULT 11	FAULT 12	FAULT 13	FAULT 14	FAULT 15	FAULT 16	FAULT 17	FAULT 18	FAULT 19	FAULT 20	DETECTION	ISOLATION
OPERATION GRAPHS	NOC	27.40	0	66.98	0	0	0	0	0	0	1.46	0	0.73	0	0.21	0.42	1.98	0.31	0.31	0.21	0	0	FalsePos	Missed
	FAULT 1	16.67	82.81	0.10	0	0	0	0	0	0	0	0	0	0.10	0.31	0	0	0	0	0	0	TruePos	Correct
	FAULT 2	32.60	0	59.06	0	0	0	0	0	0	0.31	0.31	0.21	0	0	0.42	3.96	0.21	0.31	1.35	0.73	0.52	TruePos	Correct
	FAULT 3	16.67	0	0	73.85	0	0	0	0	0	3.33	0.42	0	0	0	0.42	2.71	0	0	0	2.60	0	TruePos	Correct
	FAULT 4	16.67	0	0	5.63	0	0	0	0.10	0	1.98	2.08	0	0	0	0	0	0	0	73.54	0	0	TruePos	Missed
	FAULT 5	18.44	0	0	0	0	0	0	0	0	0.10	0.42	0.10	9.38	0	0.83	0.63	0.10	0	0.52	69.17	0.31	TruePos	Missed
	FAULT 6	16.67	0	0	0	0	77.92	0	0	0	0.10	0	0	5.31	0	0	0	0	0	0	0	0	TruePos	Missed
	FAULT 7	16.67	0	0	0	0	5.52	0.21	0.10	0	0	0	0	2.40	0	0	0	0	0	73.23	0	1.88	TruePos	Missed
	FAULT 8	59.90	0	0	0	0	0	0	0	0	0.21	0.31	0.10	0	0.10	0.21	11.56	0	0.52	1.56	25.42	0.10	FalseNeg	Missed
	FAULT 9	16.67	0	0	0	0	0	0	0	0.31	0	0	0	0.31	80.10	0.10	2.29	0	0	0.10	0.10	0	TruePos	Missed
	FAULT 10	16.88	0	3.85	0	0	0	0	0	0	2.60	1.46	0	0	0	1.46	8.54	1.25	0	0.31	63.65	0	TruePos	Missed
	FAULT 11	16.67	0	0	0	83.23	0	0	0	0	0.10	0	0	0	0	0	0	0	0	0	0	0	TruePos	Missed
	FAULT 12	16.67	0	0	0	0	0	0	0	0.10	0.10	0	0	0.52	78.65	0.10	3.65	0	0.10	0.10	0	0	TruePos	Missed
	FAULT 13	16.67	1.67	0.10	0	0	0	0	0	3.75	7.60	0.52	0	0.21	0	65.31	3.85	0	0.10	0.10	0	0.10	TruePos	Missed
	FAULT 14	16.67	0	0.10	0	0	0	0	0	24.48	11.46	0	0	0.31	43.44	0.10	3.33	0	0.10	0	0	0	TruePos	Missed
	FAULT 15	16.67	0	0.10	0	0	0	0	0	0.31	0.21	0	0	0.42	79.06	0	3.02	0	0	0.21	0	0	TruePos	Missed
	FAULT 16	16.67	0	13.75	60.63	0	0	0	0	0	0.63	3.23	0	0	0	1.88	3.13	0	0.10	0	0	0	TruePos	Missed
	FAULT 17	16.67	0	0	0	83.23	0	0	0	0	0	0	0	0	0	0	0	0.10	0	0	0	0	TruePos	Missed
	FAULT 18	16.67	0	0	7.50	0	0	0	0	0	1.25	3.65	0	0	0	0.31	1.77	0	0	68.85	0	0	TruePos	Correct
	FAULT 19	16.67	0	26.35	0	0	0	0	0	0	0.83	0.42	0	0	0	1.67	1.25	0.63	0	0	52.08	0.10	TruePos	Correct
	FAULT 20	16.67	0	0	0	34.90	0	0.10	0	0	0.10	0	0	0	0	0	0	0	0	0	0	48.23	TruePos	Correct

The mean performance results observed for the 3 Quantitative approach permutations present a similar behaviour for accuracy, where both alternative permutations delivered higher results than the default permutation. Alternatively, the default permutation presents a middle-ground for isolation, as the SHEOM MEAN demonstrated minor improvement over the default, while the IQR-HEOM MEAN delivered a reduced isolation capability in comparison to the default permutation. MCC results by both HEOM-MEAN and SHEOM-MEAN are fair, where a contrast is seen for the IQR-HEOM MEAN permutation. The mean MCC value for this permutation is indicative of the false positive classification observed in Table 5.23, which is not the only instance experienced for the range of datasets utilised.

The standard deviation for these respective performance measures is included per Table 5.25, where additional complexity is introduced when discerning which permutation demonstrated the best performance. The IQR-HEOM MEAN permutation demonstrated the smallest standard deviation for each of the respective performance measures, yet the negative MCC value delivered by this permutation still casts doubt to its capability in correctly discerning between normal and fault conditions.

Table 5.24: Performance measures - Quantitative approach

	HEOM MEAN	HEOM MEAN	HEOM MEAN
ACCURACY (%)	71.429	79.524	84.762
ISOLATION (%)	25.238	29.048	23.810
MCC	0.327	0.313	-0.075

Table 5.25: Standard deviation of performance measures - Quantitative approach

	SHEOM MEAN	SHEOM MEAN	SHEOM MEAN
ACCURACY (%)	8.781	7.982	4.666
ISOLATION (%)	6.042	6.547	3.689
MCC	0.066	0.233	0.020

5.5.2 Sensitivity analysis: Quantitative approach

The sensitivity analysis performed for the Quantitative approach presented the use of a single factor ANOVA to determine the significance of a single independent variable as cause to the variation observed. This variation is observed as between groups for the range of functional parameter permutations. Table 5.26 illustrates the first ANOVA performed to determine whether normalisation presented significant cause to the variation observed. With the F-value calculated exceeding the F-critical value, the null hypothesis stating that no significant variation exists between group means should be rejected, and the alternate hypothesis stating the existence to such significant variation exists between group means should be accepted. Additionally, the sum of squares within group variation is observed to be fairly less than between groups, indicating fair consistency between MCC values by the respective permutations.

ANOVA illustrated in Table 5.27 for the accuracy values delivered by the respective permutations delivers a similar case with the calculated F-value exceeding the F-critical values, and is indicative of the normalisation being a significant cause to variation observed between samples groups. Therefore, this null hypothesis should be rejected and the alternate hypothesis should be accepted. Note the size of the sum of squares calculated for within groups, as this comprises the majority share of the total sum of squares for this analysis.

Analysis of the isolation values presented by Table 5.28 indicates the opposite result as observed for the other performance measures, where the calculated F-value did not exceed the F-critical value. Thus, no significant cause to variation between sample groups was observed and the null hypothesis should be accepted.

Table 5.26: ANOVA MCC - Quantitative approach

Anova: Single Factor — Quantitative Approach - MCC

SUMMARY

Groups	Count	Sum	Average	Variance
HEOM	10	3.27	0.33	4.88E-03
SHEOM	10	3.13	0.31	0.06
IQR-HEOM	10	-0.75	-0.07	4.23E-04

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Between Groups	1.04	2	0.52	23.82	1.09E-06	3.35	FALSE
Within Groups	0.59	27	0.02				
Total	1.63	29					

Table 5.27: ANOVA Accuracy - Quantitative approach

Anova: Single Factor — Quantitative Approach - Accuracy

SUMMARY

Groups	Count	Sum	Average	Variance
HEOM	10	714.29	71.43	85.66
SHEOM	10	795.24	79.52	70.80
IQR-HEOM	10	847.62	84.76	24.19

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Between Groups	902.49	2	451.25	7.49	0.00	3.35	FALSE
Within Groups	1625.85	27	60.22				
Total	2528.34	29					

The sensitivity values calculated in Table 5.29 align with the hypothesis accepted for each of the respective ANOVAs performed. The selection of the normalisation value presented significant cause to variation of the MCC and the accuracy delivered by the permutations. Sensitivity of the isolation towards normalisation was determined to be small, and is consistent with the acceptance of the null hypothesis observed in Table 5.28. The remainder of sensitivity, seen in Figure 5.3, demonstrated towards within the sample groups is indicative of a moderate difference in samples delivered by the Quantitative approach permutations.

Table 5.28: ANOVA Isolation - Quantitative approach

Anova: Single Factor — Quantitative Approach - Isolation

SUMMARY

Groups	Count	Sum	Average	Variance
HEOM	10	252.38	25.24	40.56
SHEOM	10	290.48	29.05	47.62
IQR-HEOM	10	238.10	23.81	15.12

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Between Groups	146.64	2	73.32	2.13	0.14	3.35	TRUE
Within Groups	929.71	27	34.43				
Total	1076.34	29					

Table 5.29: Sensitivity values - Quantitative approach

Source of Variation	MCC	Accuracy	Isolation
Normalisation	0.638	0.357	0.136
Within	0.362	0.643	0.864

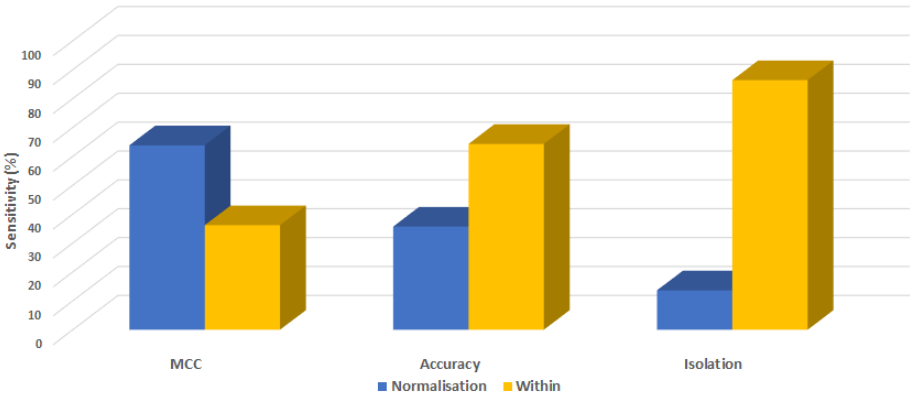

Figure 5.3: Performance measure sensitivity - Quantitative approach

5.6 Conclusion

Following the execution of both FDI and subsequent sensitivity analysis, a discernible difference could be observed in the performance results delivered by the default, as well as alternate permutations for each of the respective FDI approaches.

The Distance approach demonstrated a moderate-to-high accuracy, coupled with a fair isolation capability for the respective operation conditions observed in this study. Clear sensitivity is demonstrated towards the cost-matrix utilisation functional parameter, where a difference in performance could be observed between the functional parameter's values when implemented individually. An additional finding resulting from the implementation of multiple functional parameters is that separate permutations demonstrated the highest values for accuracy and isolation respectively. An 8% and 10% improvement was observed to the default permutation's accuracy and isolation values respectively, while standard deviation of results by these two permutations were lower than those observed for the default permutation.

The Qualitative approach demonstrated a lower performance overall in comparison to the Distance approach, where varied significance was observed towards the functional parameters analysed for the respective performance measures. Accuracy was moderate, with values ranging between 50-65%, while only being capable in correctly isolating 14-25% of conditions observed. Additionally, multiple negative MCC-values were recorded for select permutations, indicating that these are inconsistent in their ability to discern between normal and fault conditions. Sensitivity towards the functional parameters varied amongst performance measures, where both normalisation and cost-matrix utilisation was identified as significant sources to variation with high interaction noted between these parameters in the MCC values. Analysis of the accuracy values only indicated the cost-matrix utilisation in presenting a significant source of variation, while analysis of isolation values identified both functional parameters as significant. For each of the ANOVA's performed for Qualitative approach permutations, significant variation within samples was observed for the performance results delivered by these permutations. This indicates an inconsistency of their capability in discerning between conditions observed in performing FDI.

The Quantitative approach demonstrated improved accuracy compared to the Qualitative approach, while presenting isolation rates slightly below those observed for the Distance approach. While the IQR-HEOM MEAN permutation delivered a higher accuracy, it also presented negative MCC values as an indication of this permutation having incorrectly classified multiple normal conditions amongst the datasets utilised. This draws attention to being cautious in assuming performance without considering a permutation's ability to consistently discern between normal- and fault conditions. Lastly, the normalisation functional parameter proved significant to variation observed for both MCC and accuracy values, while not presenting significant to isolation performance by the respective Quantitative approach permutations.

Chapter 6

Validation of sensitivity analysis

6.1 Introduction

Following FDI and sensitivity analysis performed of the various EGBV methods in Chapter 5, the results observed required validation to determine whether the proposed concept of functional parameter sensitivity holds true under different operational conditions. Validation for this study would consist of setting the operation of the Tennessee Eastman process to Mode 5 [9]. This will alter the production output ratio (G/H) delivered by the TEP model from 50/50 to 10/90, where fault conditions could present different node signatures within the model. This chapter will follow the model setup for Mode 5, followed by performing both FDI and sensitivity analysis using a similar structure implemented in Chapter 5. This includes executing 18 Distance approach permutations, 15 Qualitative approach permutations, and 3 Quantitative approach permutations for the 10 datasets simulated for the 10/90 product ratio.

The FDI procedure was implemented identically to what was done in Chapter 5, without any deviation in methodology, to allow an accurate comparison of performance results and identifying consistent behaviour in the EGBV methods. FDI was first performed using the range of functional parameter permutations for each respective EGBV method, and accompanied by the sensitivity analysis performed similar to Chapter 5. The FDI and sensitivity analysis results for Mode 5 datasets were first discussed separately on notable behaviour demonstrated by the EGBV methods for this TEP operation mode. This discussion is presented on the comparison of results by the EGBV methods for the Base case mode and Mode 5 operating modes respectively. This comparison of results would allow the EGBV methods to be inspected with the aim of identifying whether behaviour in performance and functional parameter sensitivity held constant, or if deviation could be observed between TEP operation modes.

6.2 Purpose of validation

Validation is the substantiation that the concept possesses a satisfactory range of accuracy consistent with the intended application domain [26]. The validity of the proposed functional parameter sensitivity concept needs to be determined in respect to stated research question as to its correctness. Evaluating this "correctness" would serve to determine whether the concept would be valid for one set of experimental conditions and invalid for another [26]. The scope by which concept validity could be considered when applied under different conditions can be expanded into two separate cases:

- Performing concept validation through only using a single model under separate operating conditions would result in the validation only being proven for a specific use case. Validation of the proposed concept would then only be considered to hold true for application to the specific model implemented.
- Performing concept validation through the use of an additional/external model than used for initial experiment results would provide a broader environment in observing how the proposed concept holds true for operation conditions that are presented differently to those observed for the initial experiment. Validation of the proposed concept could then be considered true for a wider range of use, with less restrictions imposed to model conditions.

The validation of the functional parameter sensitivity concept in this study would be performed as per this first case, where the TEP would be configured to operate at a different operation mode to deliver known operating conditions differently, as previously observed for the initial experiment. Doing so would serve to establish initial confidence in the proposed concept of functional parameter sensitivity [26]. The purpose in performing this validation could then be stated as two qualitative questions by which confidence in the concept could be evaluated:

- Do EGBV method functional parameters demonstrate consistent performance behaviour between TEP operation modes?

- Do EGBV methods demonstrate consistent sensitivity towards the respective functional parameters between TEP operation modes?

Evaluating whether a model or concept is valid for use is commonly based on the accuracy required by its intended use, where an acceptable range for any deviation would be specified. When performing validation testing, results observed within this acceptable range would confirm the model or concept to be valid for its intended purpose. An alternative exists, where accuracy is instead evaluated qualitatively as when exploring a concept/model, towards gaining insight into its behaviour or performance for specific conditions observed. The

qualitative-based evaluation is suitable for this study as both performance and sensitivity behaviour would be evaluated for consistent behaviour, as opposed to a fixed acceptable range defined. Thus the concept validation would be based on whether the consistency in behaviour by functional parameters for EGBV methods is reasonable [26].

6.3 TEP Model: Mode 5 Operation

From the range of operation modes available for the TEP model, Mode 5 in [9] was selected as a suitable option for generating additional datasets to be used for the concept validation. Mode 5 consists of altering production constraints to deliver output products G/H at mass ratios 10/90 under maximum capacity for the system. Table 6.1 presents the setpoint values for both Mode 5 and the base case mode, as implemented for the initial sensitivity analysis of the EGBV functional parameters.

Table 6.1: Setpoint values for TEP operation modes [9]

Number	Measure	Unit	Base Case (50/50)	Mode 5 (10/90)
XMEAS(1)	A Feed	kscmh	0.251	0.325
XMEAS(2)	D Feed	kg/h	3664	761
XMEAS(3)	E Feed	kg/h	4509	8354
XMEAS(4)	A +C Feed	kscmh	9.35	8.87
XMEAS(5)	Recycle flow	kscmh	26.90	31.27
XMEAS(6)	Reactor feed	kscmh	42.34	46.24
XMEAS(7)	Reactor pressure	kPa	2705	2800
XMEAS(8)	Reactor level	%	75	65.0
XMEAS(9)	Reactor temperature	°C	120.4	124.6
XMEAS(10)	Purge rate	kscmh	0.337	0.384
XMEAS(11)	Sep temperature	°C	80.1	88.9
XMEAS(12)	Sep level	%	50	50.0

XMEAS(13)	Sep pressure	kPa	2634	2705
XMEAS(14)	Sep underflow	m^3/h	25.16	27.45
XMEAS(15)	Stripper level	%	50	50.0
XMEAS(16)	Stripper pressure	kPa	3102	3330
XMEAS(17)	Stripper underfow	m^3/h	22.95	23.55
XMEAS(18)	Stripper temperature	°C	65.73	63.9
XMEAS(19)	Steam flow	kg/h	230	5.11
XMEAS(20)	Compressor work	kW	341	271.7
XMEAS(21)	React. cool temperature	°C	94.6	108.5
XMEAS(22)	Condo. cool temperature	°C	77.3	89.8
XMEAS(23)	Feed - A	mol%	32.19	34.78
XMEAS(24)	Feed - B	mol%	8.89	7.85
XMEAS(25)	Feed - C	mol%	26.38	19.54
XMEAS(26)	Feed - D	mol%	6.88	1.24
XMEAS(27)	Feed - E	mol%	18.78	26.03
XMEAS(28)	Feed - F	mol%	1.66	5.60
XMEAS(29)	Purge - A	mol%	32.96	36.71
XMEAS(30)	Purge - B	mol%	13.82	11.47
XMEAS(31)	Purge - C	mol%	23.98	14.57
XMEAS(32)	Purge - D	mol%	1.26	0.13
XMEAS(33)	Purge - E	mol%	18.58	22.92
XMEAS(34)	Purge - F	mol%	2.26	7.44
XMEAS(35)	Purge - G	mol%	4.84	1.26

XMEAS(36)	Purge - H	mol%	2.30	5.51
XMEAS(37)	Product - D	mol%	0.02	0.00
XMEAS(38)	Product - E	mol%	0.84	1.01
XMEAS(39)	Product - F	mol%	0.10	0.32
XMEAS(40)	Product - G	mol%	53.72	11.65
XMEAS(41)	Product - H	mol%	43.83	85.53

With changes to the above-mentioned measurement setpoints, a near-identical structure to generating datasets was followed as implemented for the initial datasets acquired for this study. This includes using identical random seed values for datasets 1 to 10, and performing a simulation of operation conditions at the same duration as done previously. This includes 25 hours simulated for reference conditions, and 48 hours for operation conditions. Deviation from the previously implemented structure occurs with the duration of simulated time, after which fault conditions may occur. This is due to the initial startup of the model occurring at 50/50, and requires additional time for the system to transition towards the 10/90 setpoint at normal operating conditions. This initial normal operating period was extended from 1 hour, as implemented for datasets previously acquired for this study, to 8 hours to ensure the system has reached the 10/90 production ratio setpoint at normal operation, prior to introducing fault conditions to the system.

Figure 6.1 illustrates the ratio of products G/H produced during normal operating conditions for both base case mode and Mode 5 operation of the TEP model. Data demonstrated in this figure is for XMEAS(40) and XMEAS(41), simulated for both initial and validation dataset 2 (random seed 2). Mode 5 operation did not reach the G/H setpoint values as XMV1 was constrained to 20% while XMV2 was saturated at 200% during operation. This resulted in the reactant D Feed being in excess of the required SP with this material reducing capacity available for other reactant material in the reactor. This imbalance of reactant material would then have prevented optimal reactor operation and limit the output of products to reach the required amount for this operation mode. While Mode 5 was not able to reach the G/H production setpoints, it is still accepted for this study in providing significant variation to the previously observed base case mode operating conditions.

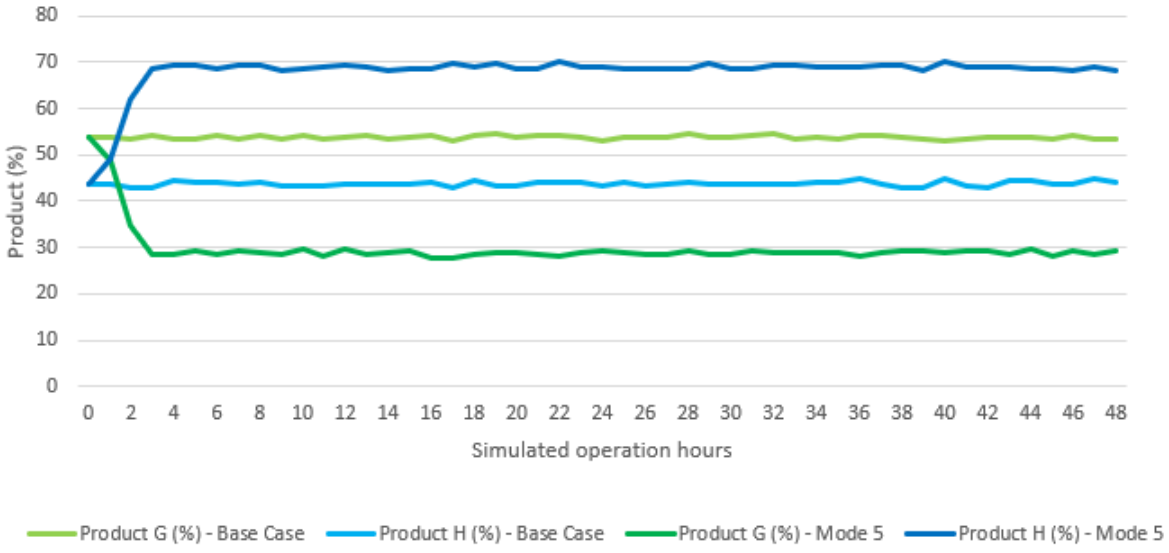

Figure 6.1: TEP model product ratios

6.4 Validation - Distance approach

In performing validation of the FDI results previously observed for the Distance approach, each of the 18 permutations were repeated using the Mode 5 datasets acquired for this chapter. The default HEOM-MEAN permutation demonstrated a moderately high average accuracy of 85% as average for the Mode 5 datasets, where accuracy by the other permutations illustrated in Table 6.2 ranged between 83.3-93.3% in close proximity to the default permutation. Permutations implementing the RC: SMALLEST 1 option, demonstrated a clear deviation by delivering the highest values, in comparison to the base case mode results where these permutations previously delivered the lowest accuracy values. The remainder of permutations demonstrate consistent behaviour to accuracy between the two respective operation modes.

Table 6.2: Mode 5 performance measures: Distance approach

	HEOM MEAN	HEOM RC: SMALLEST 1	HEOM RC: SMALLEST 5	HEOM RC: SMALLEST 10	HEOM Threshold: 2-SD MIN	HEOM Threshold: 3-SD MIN
ACCURACY (%)	85.238	93.333	87.619	85.238	87.143	83.333
ISOLATION (%)	35.714	2.857	9.048	15.238	23.333	27.143
MCC	-0.017	-0.017	-0.060	-0.014	0.055	0.077
	SHEOM MEAN	**SHEOM RC: SMALLEST 1**	**SHEOM RC: SMALLEST 5**	**SHEOM RC: SMALLEST 10**	**SHEOM Threshold: 2-SD MIN**	**SHEOM Threshold: 3-SD MIN**
ACCURACY (%)	86.190	93.333	88.095	85.238	87.143	83.333
ISOLATION (%)	37.619	2.857	9.524	17.619	26.667	30
MCC	-0.013	-0.017	-0.059	-0.017	0.055	0.072
	IQR-HEOM MEAN	**IQR-HEOM RC: SMALLEST 1**	**IQR-HEOM RC: SMALLEST 5**	**IQR-HEOM RC: SMALLEST 10**	**IQR-HEOM Threshold: 2-SD MIN**	**IQR-HEOM Threshold: 3-SD MIN**
ACCURACY (%)	89.524	92.857	87.619	84.762	89.048	92.857
ISOLATION (%)	44.286	5.714	9.524	20.476	36.667	42.381
MCC	0.172	0.047	-0.013	0.034	0.294	0.450

Isolation performance delivered for Mode 5 operation presents a more varied case in comparison to the base case mode in Chapter 5. The default HEOM-MEAN permutation demonstrates consistent behaviour by delivering 35.7% isolation, compared to the 36.6% isolation delivered for the base case mode. Significant variation is amongst the range of functional parameter permutations with isolation values ranging between 2.8-44.2%, as the majority of permutations demonstrated a lower isolation performance than the default permutation. Additionally, the individual permutations presenting the highest isolation performance for Mode 5 operation differ to those previously demonstrating highest isolation performance for the base case mode. This presents a case to isolation performance being inconsistent for the alternate permutations of the Distance approach between the respective operation modes for the TEP model.

MCC values delivered for mode 5 operation present an interesting observation, where 9 of the permutations implemented, including the default HEOM-MEAN permutation, delivered a negative mean MCC value for the Mode 5 operation datasets. This indicates that these permutations could not consistently discern the NOC from the fault conditions experienced. More consistent MCC values are observed for permutations implementing IQR-HEOM normalisation or either of the two threshold-based options to cost-matrix utilisation. The IQR-HEOM Threshold: 3-SD Min permutation delivered the highest average MCC value for Mode 5 operation, and is consistent with its MCC value of 0.424% previously observed for base case mode operation.

Table 6.3: Standard deviation of performance measures: Distance approach (Mode 5)

	HEOM MEAN	HEOM RC: SMALLEST 1	HEOM RC: SMALLEST 5	HEOM RC: SMALLEST 10	HEOM Threshold: 2-SD MIN	HEOM Threshold: 3-SD MIN
ACCURACY (%)	5.813	3.159	3.810	8.896	4.786	5.734
ISOLATION (%)	5.324	3.159	4.972	5.553	3.956	4.786
MCC	0.162	0.027	0.024	0.145	0.247	0.245
	SHEOM MEAN	**SHEOM RC: SMALLEST 1**	**SHEOM RC: SMALLEST 5**	**SHEOM RC: SMALLEST 10**	**SHEOM Threshold: 2-SD MIN**	**SHEOM Threshold: 3-SD MIN**
ACCURACY (%)	4.972	3.159	3.194	8.371	4.786	5.324
ISOLATION (%)	6.547	3.159	5.216	4.786	4.364	3.719
MCC	0.160	0.027	0.022	0.131	0.247	0.235
	IQR-HEOM MEAN	**IQR-HEOM RC: SMALLEST 1**	**IQR-HEOM RC: SMALLEST 5**	**IQR-HEOM RC: SMALLEST 10**	**IQR-HEOM Threshold: 2-SD MIN**	**IQR-HEOM Threshold: 3-SD MIN**
ACCURACY (%)	6.317	3.194	5.303	5.553	7.392	6.818
ISOLATION (%)	4.286	2.857	6.023	5.238	5.654	6.190
MCC	0.365	0.216	0.124	0.214	0.385	0.439

Standard deviation (SD) values for the respective performance measures are illustrated in Table 6.3, where the majority of permutations delivered lower SD values for both accuracy and isolation when compared to their respective values delivered for the base case mode. SD of MCC values presented higher across all permutations for Mode 5 operation than demonstrated in the base case mode previously observed in Chapter 5.

From the FDI results delivered by the Distance approach permutations, the Two-way ANOVA with replication was performed to establish the presence of functional parameter sensitivity for each of the performance measures. Table 6.4 demonstrates the ANOVA performed in analysing the variance observed amongst MCC values delivered by the respective permutations. Following the same notation for normalisation (row) and cost-matrix utilisation (column) in the ANOVA table, the following observations to the null hypotheses are made:

- The F-value calculated for normalisation (Row) is significantly higher than the associated F-critical value, indicating a significant variance that is observed between the respective normalisation options. The null hypothesis shall be rejected and the alternate hypothesis of significant difference existing between normalisation group means can be accepted.

- The F-value for cost-matrix utilisation (Column) is larger than the F-critical value and the alternate hypothesis, therefore the hypothesis that a significant difference in group means exists should be accepted.

- The F-value for interaction between functional parameters, when compared to the associated F-critical value, indicates that the null hypothesis can be accepted, as no significant difference was observed between group means.

An analysis of accuracy values was delivered by the Distance approach permutations for Mode 5 operation, as illustrated in Table 6.5, where an identical case was observed for the MCC values:

- The F-value for normalisation (Row) is larger than the F-critical value, and the null hypothesis should be rejected. The alternate hypothesis can be accepted, as a significant difference in group was observed.

- The F-value for cost-matrix utilisation (Column) is larger than the F-critical value, and the alternate hypothesis that significant difference in group means exists can be accepted.

- The F-value for interaction between functional parameters, when compared to the associated F-critical value, indicates that the null hypothesis can be accepted, as no significant difference was observed between group means.

Table 6.4: ANOVA MCC - Distance approach (Mode 5)

Anova: Two-Factor With Replication — **Distance Approach - MCC**

SUMMARY	MEAN	RC: SMALLEST 1	RC: SMALLEST 5	RC: SMALLEST 10	THRESHOLD: 2-SD MIN	THRESHOLD: 2-SD MIN	Total
HEOM							
Sum	-0.17	-0.17	-0.60	-0.14	0.55	0.77	0.24
Average	-0.02	-0.02	-0.06	-0.01	0.06	0.08	3.93E-03
Variance	0.03	8.10E-04	6.29E-04	0.02	0.07	0.07	0.03
SHEOM							
Sum	-0.13	-0.17	-0.59	-0.17	0.55	0.72	0.22
Average	-0.01	-0.02	-0.06	-0.02	0.06	0.07	3.63E-03
Variance	0.03	8.10E-04	5.36E-04	0.02	0.07	0.06	0.03
IQR-HEOM							
Sum	1.72	0.47	-0.13	0.34	2.94	4.50	9.83
Average	0.17	0.05	-0.01	0.03	0.29	0.45	0.16
Variance	0.15	0.05	0.02	0.05	0.16	0.21	0.13
Columns							
Count	30	30	30	30	30	30	
Sum	1.42	0.12	-1.32	0.04	4.04	5.98	
Average	0.05	4.05E-03	-0.04	1.28E-03	0.13	0.20	
Variance	0.07	0.02	0.01	0.03	0.11	0.14	
ANOVA							
Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***	***Null Hyphothesis***
Row	1.03	2	0.51	9.10	1.79E-04	3.05	FALSE
Column	1.28	5	0.26	4.53	6.80E-04	2.27	FALSE
Interaction	0.59	10	0.06	1.04	0.41	1.89	TRUE
Within	9.12	162	0.06				
Total	12.01	179					

Table 6.5: ANOVA Accuracy - Distance approach (Mode 5)

Anova: Two-Factor With Replication	Distance Approach - Accuracy						
SUMMARY	MEAN	RC: SMALLEST 1	RC: SMALLEST 5	RC: SMALLEST 10	THRESHOLD: 2-SD MIN	THRESHOLD: 2-SD MIN	Total
HEOM							
Sum	852.38	933.33	876.19	852.38	871.43	833.33	5219.05
Average	85.24	93.33	87.62	85.24	87.14	83.33	86.98
Variance	37.54	11.09	16.12	87.93	25.45	36.53	42.94
SHEOM							
Sum	861.90	933	880.95	852.38	871.43	833.33	5233.33
Average	86.19	93	88.10	85.24	87.14	83.33	87.22
Variance	27.46	11.09	11.34	77.85	25.45	31.49	38.04
IQR-HEOM							
Sum	895.24	928.57	876.19	847.62	890.48	928.57	5366.67
Average	89.52	92.86	87.62	84.76	89.05	92.86	89.44
Variance	44.34	11.34	31.24	34.27	60.72	51.65	43.88
Columns							
Count	30	30	30	30	30	30	
Sum	2610	2795.24	2633.33	2552.38	2633.33	2595.24	
Average	87	93.17	87.78	85.08	87.78	86.51	
Variance	37	10.45	18.27	62.14	35.47	57.99	
ANOVA							
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Row	220.96	2	110.48	3.14	0.05	3.05	FALSE
Column	1157.47	5	231.49	6.58	1.32E-05	2.27	FALSE
Interaction	514	10	51.37	1.46	0.16	1.89	TRUE
Within	5696.15	162	35.16				
Total	7588.31	179					

The ANOVA performed for isolation values by the Distance approach present a similar case as both MCC and accuracy analysis, while demonstrating significant variance between functional parameters, as observed in Table 6.6. The ANOVA of isolation values delivered by the range of permutations indicates to both normalisation and cost-matrix utilisation, which presents a significant variance to group means respectively. The null hypothesis for both parameters can be rejected and their alternate hypothesis is to be accepted based on their F-values exceeding the respective F-critical values. Additionally, significant variance was observed in the interaction between independent variables, and is indicative of inconsistency within samples by the respective permutations. The null hypothesis for interaction between independent variables (functional parameters) should be rejected, and the alternate hypothesis can be accepted.

Table 6.6: ANOVA Isolation - Distance approach (Mode 5)

Anova: Two-Factor With Replication	Distance Approach - Isolation						
SUMMARY	MEAN	RC: SMALLEST 1	RC: SMALLEST 5	RC: SMALLEST 10	THRESHOLD: 2-SD MIN	THRESHOLD: 2-SD MIN	Total
HEOM							
Sum	357.14	28.57	90.48	152.38	233.33	271.43	1133.33
Average	35.71	2.86	9.05	15.24	23.33	27.14	18.89
Variance	31.49	11.09	27.46	34.27	17.38	25.45	147.56
SHEOM							
Sum	376.19	28.57	95.24	176.19	266.67	300.00	1242.86
Average	37.62	2.86	9.52	17.62	26.67	30.00	20.71
Variance	47.62	11.09	30.23	25.45	21.16	15.37	168.97
IQR-HEOM							
Sum	442.86	57.14	95.24	204.76	366.67	423.81	1590.48
Average	44.29	5.71	9.52	20.48	36.67	42.38	26.51
Variance	20.41	9.07	40.31	30.49	35.53	42.58	269.32
Columns							
Count	30	30	30	30	30	30	
Sum	1176.19	114.29	280.95	533.33	866.67	995.24	
Average	39.21	3.81	9.37	17.78	28.89	33.17	
Variance	44.86	11.57	30.47	32.74	56.19	71.13	
ANOVA							
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Row	1898.97	2	949.48	35.87	0.00	3.05	FALSE
Column	29302.22	5	5860.44	221.41	0.00	2.27	FALSE
Interaction	974.80	10	97.48	3.68	0.00	1.89	FALSE
Within	4287.98	162	26.47				
Total	36463.97	179					

Following the ANOVA performed for each of the performance measures, the sensitivity demonstrated towards each source of variation was calculated, and is included per Table 6.7 below. The Distance approach demonstrates significant sensitivity towards within sample groups, to indicate significant difference in MCC results between individual permutations. While the alternate hypothesis for normalisation and cost-matrix utilisation were accepted, the scale of sensitivity demonstrated in each of these sources of variation is minimal in comparison to the sensitivity associated within sample groups. Sensitivity values calculated for accuracy present a similar case, where within presents the most significant cause to variation in comparison to the individual functional parameters. Notable is the normalisation sensitivity value being minimal in comparison to the cost-matrix utilisation functional parameter. Lastly, extreme sensitivity is demonstrated towards the cost-matrix utilisation functional parameter, as influence on isolation values was observed for the Distance approach, and none of the null hypotheses were accepted for this ANOVA.

Table 6.7: Validation sensitivity values - Distance approach

Source of Variation	MCC	Accuracy	Isolation
Normalisation	0.085	0.029	0.052
Cost-Matrix Utilisation	0.106	0.153	0.804
Interaction	0.049	0.068	0.027
Within	0.760	0.751	0.118

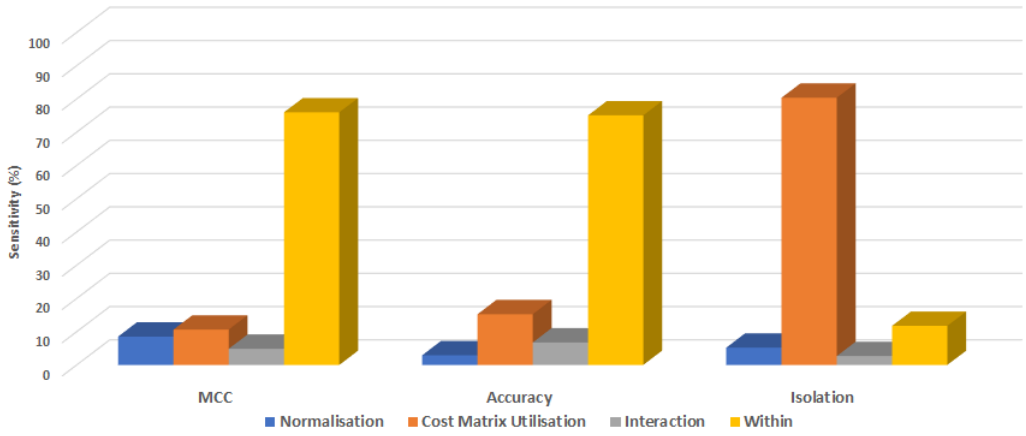

Figure 6.2: Performance sensitivity - Distance approach (Mode 5)

6.5 Validation - Qualitative approach

The Qualitative approach presented a significant overall improvement to its accuracy values delivered for Mode 5 operation, as compared to the base case mode. In Table 6.8, the default HEOM-MEAN permutation presents a 25% increase for this operation mode while maintaining a similar isolation value, as previously demonstrated for the base case mode datasets. This observation holds true for the majority of permutations, where only the HEOM 3-SD and HEOM 2-SD permutations delivered notably lower isolation values, as compared to their respective base case mode isolation values.

The default permutation for the Qualitative approach presents a similar case to its average MCC value delivered for Mode 5 operation, as was observed for the Distance approach default permutation in the previous section. Overall, the Qualitative approach demonstrated low MCC values, with only a select few permutations delivering an MCC value above 0.1. This indicates a poor consistency in discerning the NOC from fault conditions observed by this EGBV method at this operation mode. SD values illustrated in Table 6.9 indicate there is a consistency in accuracy and isolation values delivered by the Qualitative approach permutations when comparing their respective SD values between operation modes. For Mode 5 operation, the Qualitative approach permutations demonstrated larger deviation to their respective MCC values delivered in comparison to base case mode operation. With the majority of MCC SD values in Table 6.9 being larger than their respective MCC values in Table 6.8, it is indicative of the Qualitative approach permutations experiencing consistency in not being able to discern NOC from faulty conditions for Mode 5 operation.

Table 6.8: Performance measures: Qualitative approach (Mode 5)

	HEOM 3-SD	HEOM 2-SD	HEOM 1-SD	HEOM 2-MAD	HEOM 1-MAD
ACCURACY (%)	75.714	79.048	77.619	75.238	80.000
ISOLATION (%)	13.333	15.238	20.476	24.762	22.857
MCC	-0.056	-0.094	-0.007	0.127	0.162
	SHEOM 3-SD	**SHEOM 2-SD**	**SHEOM 1-SD**	**SHEOM 2-MAD**	**SHEOM 1-MAD**
ACCURACY (%)	79.048	82.857	79.524	79.048	76.667
ISOLATION (%)	21.905	24.762	25.714	28.095	25.714
MCC	-0.043	0.034	0.130	0.063	0.028
	IQR-HEOM 3-SD	**IQR-HEOM 2-SD**	**IQR-HEOM 1-SD**	**IQR-HEOM 2-MAD**	**IQR-HEOM 1-MAD**
ACCURACY (%)	84.286	83.810	81.905	81.905	73.333
ISOLATION (%)	4.286	6.667	19.524	24.762	25.714
MCC	-0.023	-0.018	0.013	0.022	0.060

ANOVA of MCC values in Table 6.10 delivered by Qualitative approach permutations present a clear case to the absence of significant difference amongst MCC results. The F-values for normalisation (Row), cost-matrix utilisation (Column), and interaction between the functional parameters (independent variables) were all less than their respective F-critical value. This confirms that no significant variance was observed between sample means, and the null hypothesis can be accepted for each of these potential causes to variance amongst sample means. This indicates that the MCC values by Qualitative approach permutations were consistent between sample groups for Mode 5 operation of the TEP model.

Table 6.9: Standard deviation of performance measures: Qualitative approach (Mode 5)

	HEOM 3-SD	HEOM 2-SD	HEOM 1-SD	HEOM 2-MAD	HEOM 1-MAD
ACCURACY (%)	9.630	8.571	8.793	13.093	9.234
ISOLATION (%)	5.948	4.666	6.406	6.999	7.619
MCC	0.175	0.039	0.181	0.232	0.260
	SHEOM 3-SD	**SHEOM 2-SD**	**SHEOM 1-SD**	**SHEOM 2-MAD**	**SHEOM 1-MAD**
ACCURACY (%)	8.025	6.459	10.444	9.085	8.371
ISOLATION (%)	5.714	6.317	7.737	8.637	4.364
MCC	0.171	0.238	0.296	0.261	0.191
	IQR-HEOM 3-SD	**IQR-HEOM 2-SD**	**IQR-HEOM 1-SD**	**IQR-HEOM 2-MAD**	**IQR-HEOM 1-MAD**
ACCURACY (%)	6.751	6.459	6.667	9.476	9.571
ISOLATION (%)	1.429	3.159	7.206	8.465	3.810
MCC	0.163	0.190	0.200	0.215	0.212

The ANOVA of accuracy values performed in Table 6.11 demonstrates a similarity to the ANOVA of MCC values. No significant variance in mean values is observed between samples groups, and the respective null hypotheses can be accepted for the analysis of accuracy values delivered by Qualitative approach permutations. An analysis of isolation values in Table 6.12 presents a contrasting case to the previous performance measures, where none of the null hypotheses can be accepted. Significant variation was observed in the sample group means associated with the normalisation, cost-matrix utilisation, interaction between functional parameters (independent variables), and variance observed within samples per group. By only accepting alternate hypotheses for analysis of isolation values, this is indicative of the Quantitative approach permutations exhibiting an inconsistent ability in isolating operation conditions at Mode 5 operation of the TEP model.

Table 6.10: ANOVA MCC - Qualitative approach (Mode 5)

Anova: Two-Factor With Replication	Qualitative Approach - MCC						
SUMMARY	3-SD	2-SD	1-SD	2-MAD	1-MAD	Total	
HEOM							
Sum	-0.56	-0.94	-0.07	1.27	1.62	1.33	
Average	-0.06	-0.09	-0.01	0.13	0.16	0.03	
Variance	3.42E-02	1.72E-03	3.62E-02	6.00E-02	7.54E-02	4.85E-02	
SHEOM							
Sum	-0.43	0.34	1.30	0.63	0.28	2.11	
Average	-0.04	3.38E-02	0.13	0.06	0.03	0.04	
Variance	3.23E-02	0.06	0.10	0.08	0.04	0.06	
IQR-HEOM							
Sum	-0.23	-0.18	0.13	0.22	0.60	0.55	
Average	-0.02	-0.02	0.01	0.02	0.06	0.01	
Variance	2.97E-02	0.04	4.44E-02	5.13E-02	4.99E-02	0.04	
Columns							
Count	30	30	30	30	30		
Sum	-1.22	-0.78	1.36	2.12	2.50		
Average	-0.04	-0.03	0.05	0.07	0.08		
Variance	0.03	0.04	0.06	0.06	0.05		
ANOVA							
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Row	0.02	2	0.01	0.25	7.79E-01	3.06	TRUE
Column	0.38	4	0.10	1.97	1.02E-01	2.44	TRUE
Interaction	0.33	8	0.04	0.84	5.72E-01	2.01	TRUE
Within	6.59	135	0.05				
Total	7.32	149					

Table 6.11: ANOVA Accuracy - Qualitative approach (Mode 5)

Anova: Two-Factor With Replication	Qualitative Approach - Accuracy						
SUMMARY	3-SD	2-SD	1-SD	2-MAD	1-MAD	Total	
HEOM							
Sum	757.14	790.48	776.19	752.38	800.00	3876.19	
Average	75.71	79.05	77.62	75.24	80.00	77.52	
Variance	103.05	81.63	85.92	190.48	94.73	105.55	
SHEOM							
Sum	790.48	828.57	795.24	790.48	766.67	3971.43	
Average	79.05	82.86	79.52	79.05	76.67	79.43	
Variance	71.55	46.36	121.19	91.71	77.85	79.08	
IQR-HEOM							
Sum	842.86	838.10	819.05	819.05	733.33	4052.38	
Average	84.29	83.81	81.90	81.90	73.33	81.05	
Variance	50.64	46.36	49.38	99.77	101.79	80.05	
Columns							
Count	30	30	30	30	30		
Sum	2390.48	2457.14	2390.48	2361.90	2300.00		
Average	79.68	81.90	79.68	78.73	76.67		
Variance	82.78	58.49	82.78	126.25	92.81		
ANOVA							
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Row	311.11	2	155.56	1.78	0.17	3.06	TRUE
Column	429.93	4	107.48	1.23	0.30	2.44	TRUE
Interaction	727.44	8	90.93	1.04	0.41	2.01	TRUE
Within	11811.79	135	87.49				
Total	13280.27	149					

Table 6.12: ANOVA Isolation - Qualitative approach (Mode 5)

Anova: Two-Factor With Replication	Qualitative Approach - Isolation						
SUMMARY	3-SD	2-SD	1-SD	2-MAD	1-MAD	Total	
HEOM							
Sum	133.33	152.38	204.76	247.62	228.57	966.67	
Average	13.33	15.24	20.48	24.76	22.86	19.33	
Variance	39.30	24.19	45.60	54.42	64.50	61.47	
SHEOM							
Sum	219.05	247.62	257.14	280.95	257.14	1261.90	
Average	21.90	24.76	25.71	28.10	25.71	25.24	
Variance	36.28	44.34	66.52	82.89	21.16	50.21	
IQR-HEOM							
Sum	42.86	66.67	195.24	247.62	257.14	809.52	
Average	4.29	6.67	19.52	24.76	25.71	16.19	
Variance	2.27	11.09	57.70	79.62	16.12	113.84	
Columns							
Count	30	30	30	30	30		
Sum	395.24	466.67	657.14	776.19	742.86		
Average	13.17	15.56	21.90	25.87	24.76		
Variance	77.70	81.22	60.36	69.88	33.47		
ANOVA							
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Row	2110.05	2	1055.03	24.50	8.38E-10	3.06	FALSE
Column	3804.38	4	951.10	22.08	4.84E-14	2.44	FALSE
Interaction	1431.90	8	178.99	4.16	1.82E-04	2.01	FALSE
Within	5814.06	135	43.07				
Total	13160.39	149					

Table 6.13 presents the sensitivity values calculated for each of the ANOVAs performed for the Qualitative approach in this chapter. For both MCC and accuracy, only significant sensitivity is observed within sample groups as indicator to inconsistent values between datasets, as opposed to a significant difference in performance between individual permutations for these measures. Figure 6.3 presents a visual illustration to the scale of sensitivity demonstrated for each of the respective performance measures. The distribution of sensitivity was more widely spread across the sources of variation for isolation values, with notable sensitivity demonstrated to both normalisation and cost-matrix utilisation. The majority share of sensitivity was still demonstrated in favour of within sample groups, for isolation values delivered by the Qualitative approach permutations.

Table 6.13: Validation sensitivity values - Qualitative approach

Source of Variation	MCC	Accuracy	Isolation
Normalisation	0.003	0.023	0.160
Cost-Matrix Utilisation	0.053	0.032	0.289
Interaction	0.045	0.055	0.109
Within	0.900	0.889	0.442

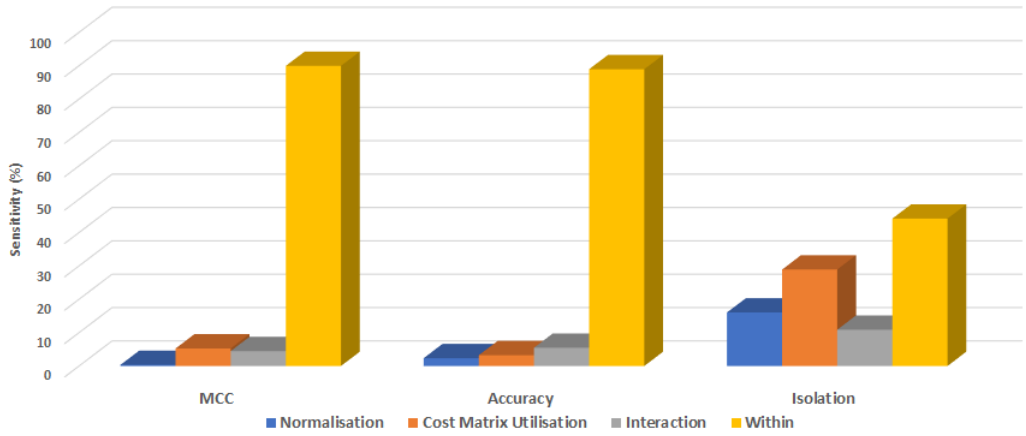

Figure 6.3: Performance sensitivity - Qualitative approach (Mode 5)

6.6 Validation - Quantitative approach

The Quantitative approach permutations delivered FDI results for the Mode 5 datasets that are consistent with those previously observed for the base case mode in Chapter 5. The accuracy values observed in Table 6.14 follow a similar trend of increasing with each permutation from left to right, as previously noted in Table 5.24 above. These accuracy values are moderately high, with the accuracy delivered by the IQR-HEOM MEAN permutation being the highest amongst all EGBV method permutations executed using the Mode 5 datasets in this chapter. Isolation values delivered are consistent with results for the base case mode datasets previously observed in Chapter 5, and present an average 2% deviation for each permutation between these operation modes. Notable deviation in MCC values could be observed between the Mode 5 and base case mode results, where MCC values in Table 6.14 were significantly lower for each permutation. The SD values illustrated in Table 6.15 present varied results when compared to the associated SD values in Chapter 5:

- HEOM-MEAN presents higher SD for accuracy, isolation and MCC for Mode 5 operation.

- SHEOM-MEAN presents higher SD for isolation while demonstrating lower SD values for both accuracy and MCC for Mode 5 operation.

- IQR-HEOM MEAN presented higher SD for isolation and MMC while having lower SD for accuracy at Mode 5 operation.

Table 6.14: Performance measures - Quantitative approach (Mode 5)

	HEOM MEAN	HEOM MEAN	HEOM MEAN
ACCURACY (%)	86.190	87.619	94.762
ISOLATION (%)	23.333	27.619	21.905
MCC	0.106	-0.056	0.064

Table 6.15: Standard deviation of performance measures - Quantitative approach (Mode 5)

	SHEOM MEAN	SHEOM MEAN	SHEOM MEAN
ACCURACY (%)	9.147	6.459	1.429
ISOLATION (%)	7.206	6.667	4.364
MCC	0.377	0.035	0.209

The ANOVAs performed on results delivered by the Quantitative approach permutations at Mode 5 operation, demonstrated only a minor deviation to the associated analysis in Chapter 5. Analysis of MCC values demonstrate that no significant variation was observed between group means, and the null hypothesis can be accepted for MCC performance. Table 6.17 which illustrates the analysis of accuracy values, presents a similar outcome as observed in Chapter 5, with the calculated F-value being larger than the F-critical value for this size table. Therefore, the null hypothesis should be rejected and the alternate hypothesis should be accepted, as significant variation was observed between sample group means. ANOVA of the isolation values in Table 6.18 also present a similar case to the associated analysis in Chapter 5, where no significant variation to sample means was observed, and the null hypothesis can be accepted.

Table 6.16: ANOVA MCC - Quantitative approach (Mode 5)

Anova: Single Factor — **Quantitative Approach - MCC**

SUMMARY

Groups	Count	Sum	Average	Variance
HEOM	10	1.06	0.11	1.58E-01
SHEOM	10	-0.56	-0.06	0.00
IQR-HEOM	10	0.64	0.06	4.85E-02

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Between Groups	0.14	2	0.07	1.02	3.74E-01	3.35	TRUE
Within Groups	1.87	27	0.07				
Total	2.01	29					

Table 6.17: ANOVA Accuracy - Quantitative approach (Mode 5)

Anova: Single Factor — **Quantitative Approach - Accuracy**

SUMMARY

Groups	Count	Sum	Average	Variance
HEOM	10	861.90	86.19	92.97
SHEOM	10	876.19	87.62	46.36
IQR-HEOM	10	947.62	94.76	2.27

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Between Groups	421.77	2	210.88	4.47	0.02	3.35	FALSE
Within Groups	1274.38	27	47.20				
Total	1696.15	29					

Sensitivity values calculated from each of these ANOVAs indicate that the majority share of sensitivity was presented towards within sample groups instead of the normalisation functional parameter. This high sensitivity to within samples, presented in Table 6.19, could be attributed to variation observed across rows and columns in the ANOVA tables, where clear a difference exists between respective permutations, and the sample values per column span a significant range, indicating inconsistent behaviour to FDI performance for Mode 5 oper-

ation. Figure 6.4 presents a visual comparison of sensitivity towards within samples, with normalisation as sole functional parameter implemented.

Table 6.18: ANOVA Isolation - Quantitative approach (Mode 5)

Anova: Single Factor — **Quantitative Approach - Isolation**

SUMMARY

Groups	Count	Sum	Average	Variance
HEOM	10	233.33	23.33	57.70
SHEOM	10	276.19	27.62	49.38
IQR-HEOM	10	219.05	21.90	21.16

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	*Null Hyphothesis*
Between Groups	176.87	2	88.44	2.07	0.15	3.35	TRUE
Within Groups	1154.20	27	42.75				
Total	1331.07	29					

Table 6.19: Validation sensitivity values - Quantitative approach

Source of Variation	MCC	Accuracy	Isolation
Normalisation	0.070	0.249	0.133
Within	0.930	0.751	0.867

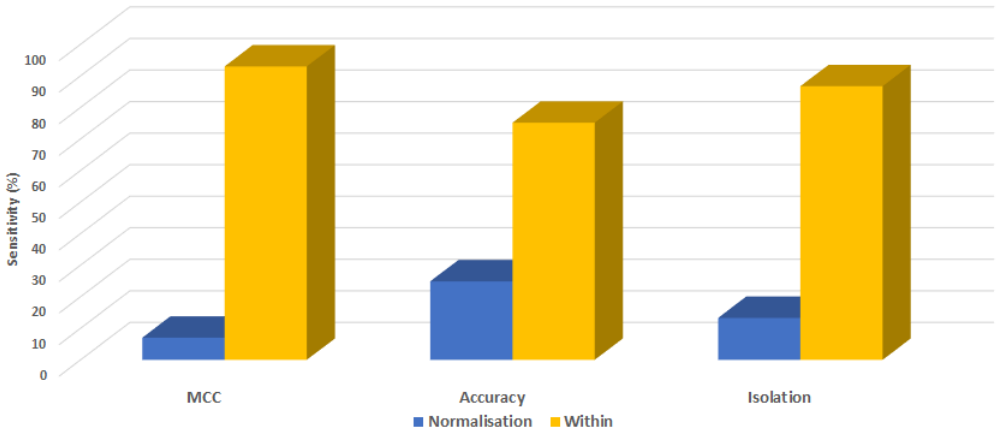

Figure 6.4: Performance sensitivity - Quantitative approach (Mode 5)

6.7 Results comparison

The FDI and sensitivity analysis performed for Mode 5 operation of the TEP model present a notable consistency to accuracy for permutations of each EGBV method, while a more significant variation is observed for isolation and MCC values delivered. This section provides a direct comparison of results between the base case mode results in Chapter 5 and results from Mode 5 that were calculated in the previous sections of this chapter.

6.7.1 Distance approach results

A comparison of the FDI and sensitivity analysis results for the Distance approach permutations presents key information to establishing consistency in performance and sensitivity demonstrated towards functional parameters. A comparison of MCC values delivered by the Distance approach permutations is presented in Figure 6.5, where results vary for nearly all of the respective permutations, with only the two IQR-HEOM threshold-based permutations proving an exception. Only the IQR-HEOM Threshold: 2-SD MIN and IQR-HEOM Threshold: 3-SD MIN presented a consistent ability to correctly discern the NOC while delivering a high accuracy for both operation modes.

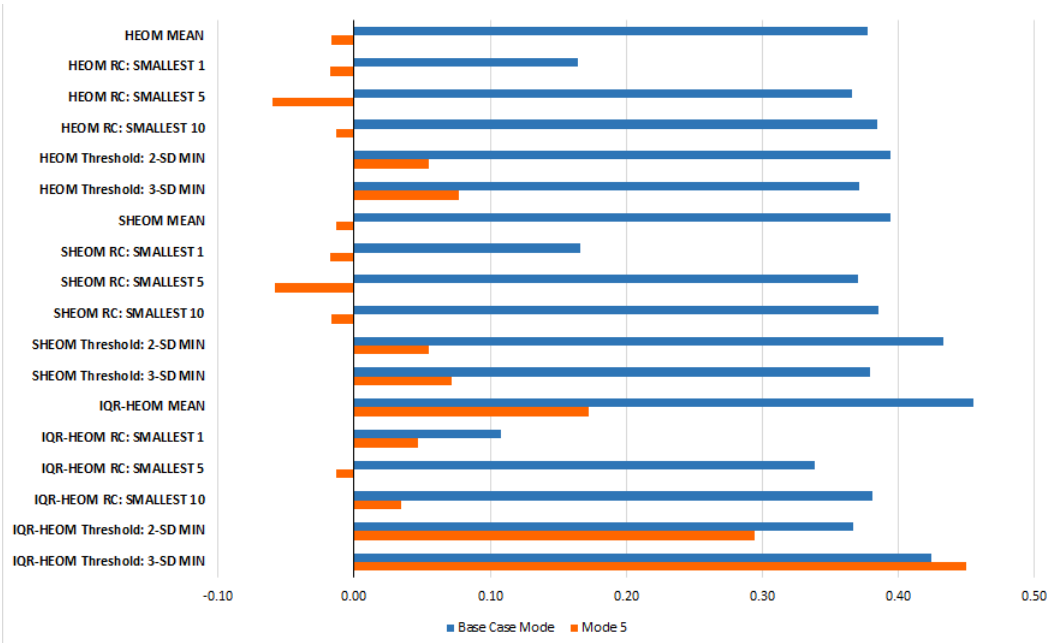

Figure 6.5: Comparison of MCC results - Distance approach

In comparing the sensitivity in MCC values demonstrated towards the respective functional parameters and their interaction, Table 6.20 presents an illustration to individual sensitivity values and the absolute delta between operation modes. Evaluating the respective ANOVA hypotheses accepted for each of the operation modes provided a basis for the following observations to be made:

- Sensitivity is demonstrated towards normalisation, as the type of hypothesis accepted differs between the operation modes investigated.

- Significant sensitivity is demonstrated towards cost-matrix utilisation, as the alternate hypothesis was accepted for both of the operation modes investigated.

- No sensitivity is demonstrated towards the interaction between normalisation and cost-matrix utilisation, as the null hypothesis was accepted for both operation modes investigated.

Table 6.20: MCC sensitivity values - Distance approach

Source	**Base Case Mode**	**Mode 5**	**Δ**
Normalisation	0.002	0.085	0.084
Cost-Matrix Utilisation	0.529	0.106	0.423
Interaction	0.033	0.049	0.015
Within	0.436	0.760	0.323

Distance approach permutations demonstrated consistent behaviour to their accuracy performance, with a majority of the results spanning between 70-95%, where a clear outlier can be observed. Permutations implementing the RC: SMALLEST 1 cost-matrix utilisation option delivered consistent lower values for the base case mode, while also some of the highest accuracy values for Mode 5 operation. While high accuracy is desirable of an FDI method, its ability to remain consistent under different operation conditions is a crucial aspect that should not be overlooked.

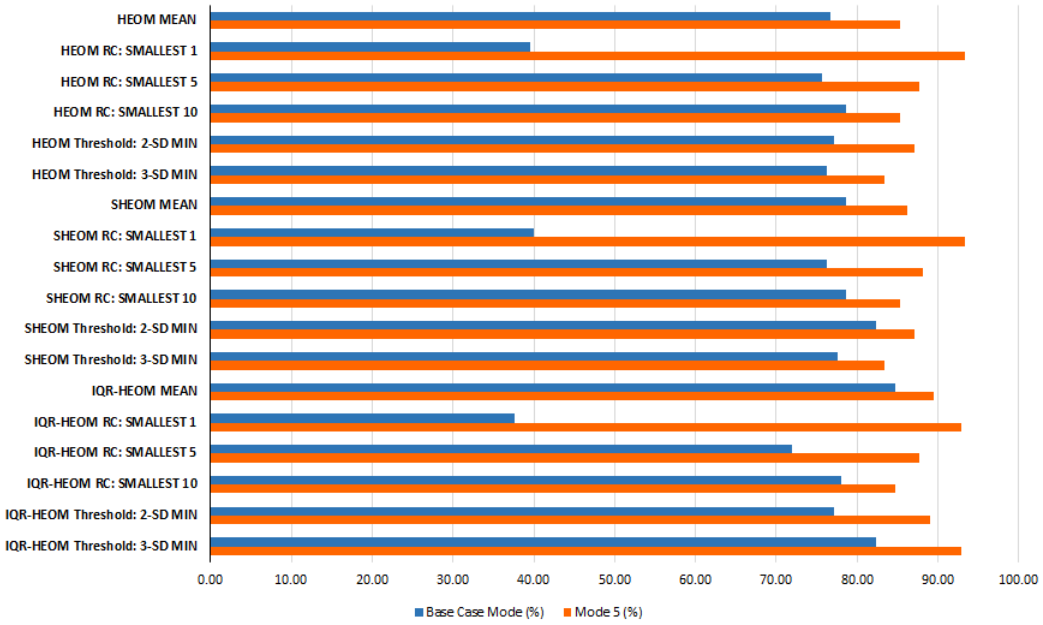

Figure 6.6: Comparison of accuracy results - Distance approach

Table 6.21 presents an illustration to individual sensitivity values and the absolute delta between operation modes, in comparing the sensitivity in accuracy values demonstrated towards the respective functional parameters and their interaction. Evaluating the respective ANOVA hypotheses accepted for each of the operation modes provide a basis for the following observations to be made:

- Sensitivity is demonstrated towards normalisation, as the type of hypothesis accepted differs between the operation modes investigated.

- Significant sensitivity is demonstrated towards cost-matrix utilisation, as the alternate hypothesis was accepted for both of the operation modes investigated.

- No sensitivity demonstrated towards interaction between normalisation and cost-matrix utilisation, as the null hypothesis was accepted for both operation modes investigated.

Table 6.21: Accuracy sensitivity values - Distance approach

Source	**Base Case Mode**	**Mode 5**	**Δ**
Normalisation	0.002	0.029	0.027
Cost-Matrix Utilisation	0.770	0.153	0.618
Interaction	0.016	0.068	0.052
Within	0.212	0.751	0.539

The isolation results delivered by the Distance approach permutations present a more distinct case, to which permutations were able to maintain consistent performance between base case mode operation and Mode 5 operation. Figure 6.7 presents the comparison of isolation values, where a noticeable trend is observed to the majority of permutations, demonstrating a clear difference in their ability to correctly isolate conditions between the respective operation modes. Permutations demonstrating clear consistency in their isolation capability include:

- HEOM-MEAN (Default)
- SHEOM-MEAN
- IQR-HEOM MEAN
- IQR-HEOM Threshold: 2-SD MIN
- IQR-HEOM Threshold: 3-SD MIN

Evaluating the respective hypotheses accepted during analysis of the isolation values for each of the operation modes provides a case for the following observations to be made:

- Sensitivity is demonstrated towards normalisation, as the type of hypothesis accepted differs between the operation modes investigated.

- Significant sensitivity is demonstrated towards cost-matrix utilisation, as the alternate hypothesis was accepted for both of the operation modes investigated.

- Sensitivity is demonstrated towards interaction between normalisation and cost-matrix utilisation, as the alternate hypothesis was accepted in the analysis of Mode 5 operation results, while the null hypothesis was accepted for the base case mode.

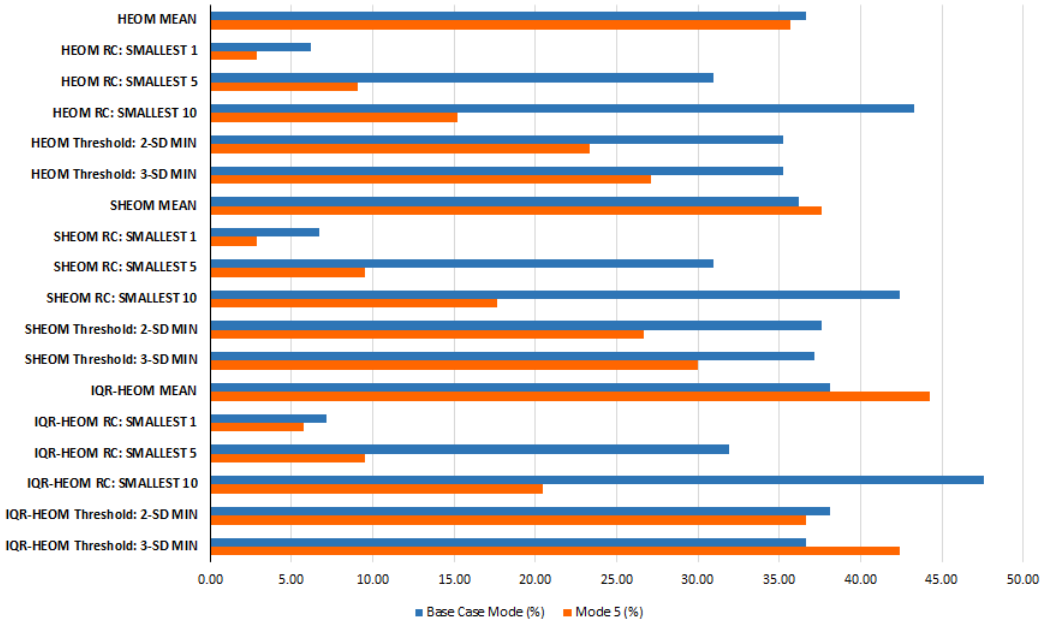

Figure 6.7: Comparison of isolation results - Distance approach

Table 6.22: Isolation sensitivity values - Distance approach

Source	**Base Case Mode**	**Mode 5**	**Δ**
Normalisation	0.004	0.052	0.048
Cost-Matrix Utilisation	0.796	0.804	0.007
Interaction	0.004	0.027	0.023
Within	0.196	0.118	0.078

6.7.2 Qualitative approach results

FDI results delivered by the Qualitative approach presents a clear difference in performance for the range permutations considered. MCC performance illustrated in Figure 6.8 presents the values delivered for the base case mode and Mode 5, with a clear indication to the consistency delivered by each permutation. Three permutations demonstrated fair consistency in their MCC values delivered for both operation modes, in comparison to the other permutations which demonstrated a high variation to MCC values between operation modes. Comparing sensitivity values in Table 6.23 and evaluating the hypotheses accepted for each of the operation modes present a case for the following observations:

- Sensitivity is demonstrated towards normalisation, as the type of hypothesis accepted differs between the operation modes investigated.

- Sensitivity is demonstrated towards cost-matrix utilisation, as the type of hypothesis accepted differs between the operation modes investigated.

- Sensitivity is demonstrated by interaction between normalisation and cost-matrix utilisation, as the type of hypothesis accepted differs between the operation modes investigated.

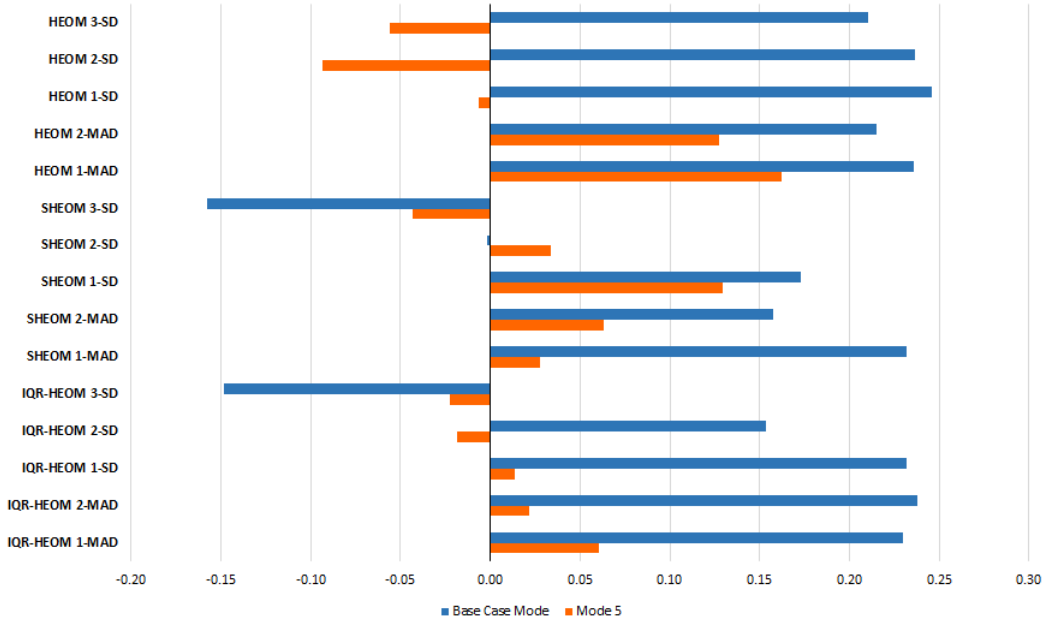

Figure 6.8: Comparison of MCC results - Qualitative approach

Table 6.23: MCC sensitivity values - Qualitative approach

Source	**Base Case Mode**	**Mode 5**	**Δ**
Normalisation	0.120	0.003	0.117
Cost-Matrix Utilisation	0.309	0.053	0.256
Interaction	0.147	0.045	0.103
Within	0.424	0.900	0.476

The accuracy performance by the Quantitative approach permutations confirms a notable consistency demonstrated for each individual operation mode, where a clear improvement to overall accuracy is demonstrated in the results for Mode 5 operation. All permutations performed demonstrated a similar level of consistency to accuracy delivered, where only minor improvement is observed between select permutations depending on the operation mode considered. Sensitivity values calculated from the ANOVA performed on the accuracy values for each of the operation modes are presented in Table 6.24, accompanied by the absolute delta in sensitivity observed for these operation modes.

A comparison is made in Table 6.24 of the sensitivity associated with the accuracy values delivered by the respective functional parameters and their interaction, to illustrate the individual sensitivity values and the absolute delta between operation modes. Evaluating the respective ANOVA hypotheses accepted for each of the operation modes provides a basis for the following observations:

- Sensitivity is demonstrated towards normalisation, as the type of hypothesis accepted differs between the operation modes investigated.

- No sensitivity is demonstrated towards cost-matrix utilisation, as the null hypothesis was accepted for both of the operation modes investigated.

- No sensitivity is demonstrated towards the interaction between normalisation and cost-matrix utilisation, as the null hypothesis was accepted for both operation modes investigated.

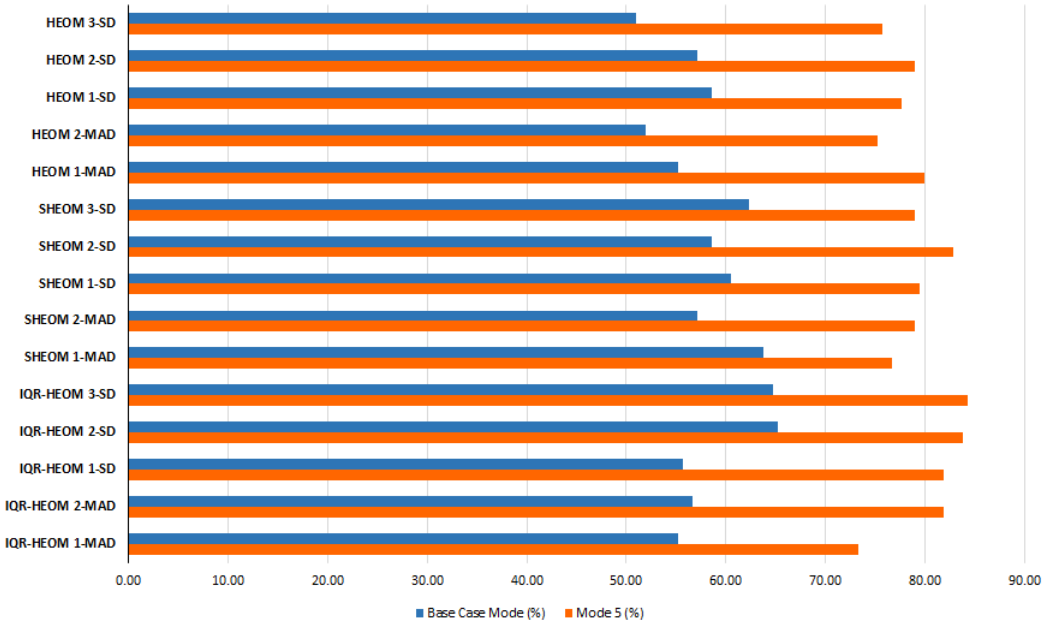

Figure 6.9: Comparison of accuracy results - Qualitative approach

Table 6.24: Accuracy sensitivity values - Qualitative approach

Source	**Base Case Mode**	**Mode 5**	Δ
Normalisation	0.064	0.023	0.040
Cost-Matrix Utilisation	0.030	0.032	0.003
Interaction	0.088	0.055	0.033
Within	0.819	0.889	0.071

The isolation performance demonstrated by the Qualitative approach permutations presents a high consistency, while the respective isolation values are low for the range defined. The default HEOM 3-SD permutation presents a consistent isolation rate of 13%, where other permutations present similar consistency at a higher isolation rate. This includes isolation results observed in Figure 6.10 for the following permutations, where near similar isolation values were delivered for both operation modes:

- HEOM 1-SD
- SHEOM 2-SD
- IQR-HEOM 2-MAD

A comparison of the sensitivity in isolation values demonstrated towards the respective functional parameters and their interaction in Table 6.25, presents an illustration of individual sensitivity values and the absolute delta in sensitivity between operation modes. Evaluating the respective ANOVA hypotheses accepted for each of the operation modes provides substantiation for the following observations made:

- Sensitivity is demonstrated towards normalisation, as the alternate hypothesis was accepted for both operation modes investigated.

- Sensitivity is demonstrated towards cost-matrix utilisation, as the alternate hypothesis was accepted for both operation modes investigated.

- Sensitivity is observed towards the interaction between normalisation and cost-matrix utilisation, as the null hypothesis was accepted for the base case mode while the alternate hypothesis was accepted for Mode 5 operation.

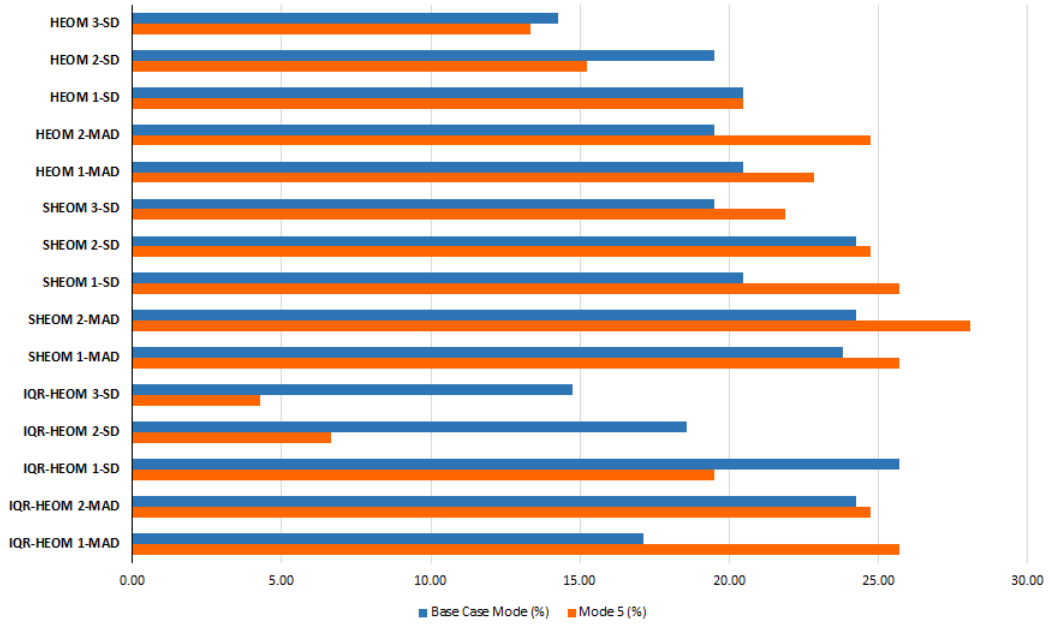

Figure 6.10: Comparison of isolation results - Qualitative approach

Table 6.25: Isolation sensitivity values - Qualitative approach

Source	**Base Case Mode**	**Mode 5**	Δ
Normalisation	0.050	0.160	0.111
Cost-Matrix Utilisation	0.117	0.289	0.172
Interaction	0.084	0.109	0.024
Within	0.749	0.442	0.307

6.7.3 Quantitative approach results

The following are FDI results observed for the Quantitative approach in the case where consistency was demonstrated in the accuracy and isolation values produced by each of the implemented permutations. The MCC values presented in Figure 6.11 demonstrate that only the default HEOM-MEAN permutation was able to consistently discern the NOC from fault conditions, observed in both operation modes investigated. The accuracy values in Figure 6.12 demonstrates a moderate consistency from the permutations, as delivered for operation at both base case mode and Mode 5. Isolation results from the Quantitative approach permutations in Figure 6.13 concur to consistent performance from the respective permutations for both operation modes investigated.

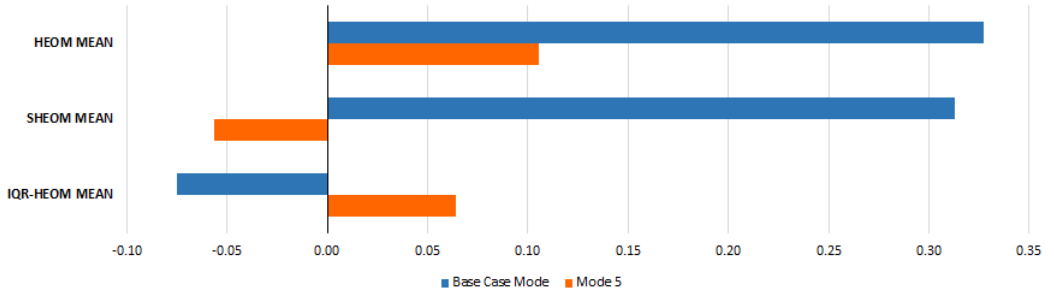

Figure 6.11: Comparison of MCC results - Quantitative approach

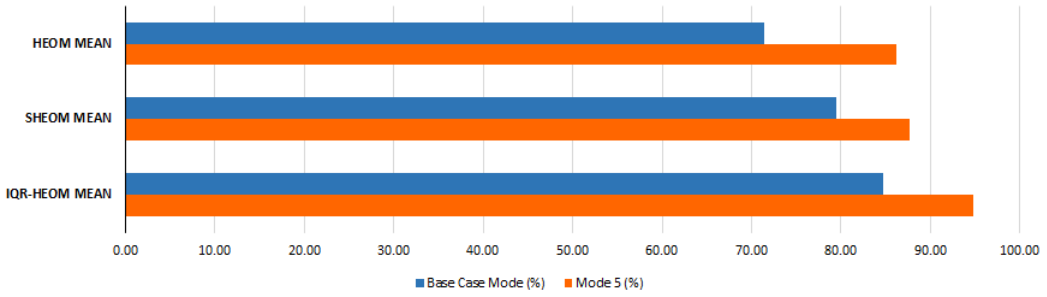

Figure 6.12: Comparison of accuracy results - Quantitative approach

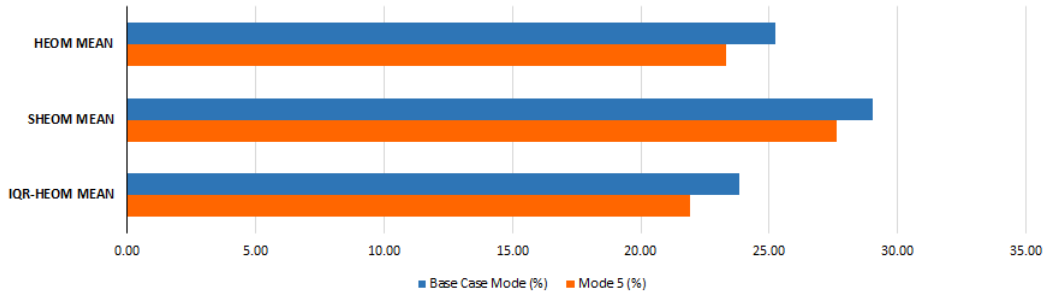

Figure 6.13: Comparison of isolation results - Quantitative approach

From the hypothesis accepted for the ANOVA performed for each respective performance measures, the following observations can be made to the sensitivity demonstrated by the Quantitative approach:

- The Quantitative approach demonstrates sensitivity to the normalisation functional parameter, as significant variation was observed for the MCC values becuase two different hypotheses were accepted for the operation modes investigated. The alternate hypothesis was accepted for the base case mode, whereas the null hypothesis was accepted for Mode 5 operation. Thus, the presence of sensitivity is accepted to be true as a result to variation still being observed for one of the operation modes investigated.

- The accuracy of the Quantitative approach does present sensitivity to the normalisation functional parameter, as the alternate hypothesis was accepted for both operation modes investigated.

- Isolation delivered by the Quantitative approach does not present sensitivity towards the normalisation functional parameter, as the null hypothesis was accepted for both operation modes investigated.

6.8 Conclusion

The validation performed in this chapter to confirm whether the proposed concept of functional parameter sensitivity holds true under different operations of the same model, delivered credible results. By executing the FDI and sensitivity analysis using operation Mode 5 datasets, the functional parameter permutations for each of the EGBV methods were tested under different operational conditions, to establish whether previously observed FDI performance and sensitivity results were consistent to those observed in Chapter 5. Comparing these FDI results between operation modes demonstrated a similarity in overall performance amongst permutations for each of the EGBV methods, while the best performing permutation for each EGBV method differed between the operation modes. This is an indication that care must be taken in selecting functional parameter options for each of the EGBV methods, as only a select few permutations demonstrated consistent performance for each of the performance measures. From the sensitivity analysis performed in this chapter, previous sources of sensitivity were either substantiated or demonstrated to the initial hypothesis not holding true for a different system operation mode. These cases were accepted to demonstrate sensitivity to the respective source of variation to sample values, as this proves to some case of sensitivity being subject to operation conditions observed by an EGBV method.

Chapter 7

Conclusion

7.1 Introduction

Functional parameters were identified for select EGBV methods and investigated to determine if these present significance to FDI performance and sensitivity in the results observed. From the FDI results delivered by the range of functional parameter permutations for each EGBV method, a clear comparison could be drawn between the default method permutation and the newly identified permutations to any deviation observed. Such deviation demonstrated as either improvement or reduced performance in each of the respective performance measures when compared to the default permutation of an EGBV method. The sensitivity analysis results observed for the initial base case mode of operation for the TEP model indicated initial significant sources of sensitivity present in each EGBV method. The validation performed in duplicating the sensitivity analysis at Mode 5 operation for the TEP model provided confirmation on some of the analysis hypotheses initially accepted, while delivering alternate findings to sensitivity demonstrated by select sources at this operation mode. Overall, the investigation to functional parameter sensitivity was successful, as new information was gained to the performance, consistency in results delivered, and sensitivity demonstrated to the functional parameters considered for this study.

7.2 Reflect on research objectives

7.2.1 Identification of relevant EGBV functional parameters

Functional parameter locations were identified for the Distance, Qualitative and Quantitative approaches, as each of these present unique steps within their respective methodology. Initial functional parameter locations identified for these EGBV methods included distance function, normalisation and cost-matrix utilisation. However, these proved to cause limitations in investigating the distance function for this study. The normalisation and cost-matrix utilisation options identified for each of these EGBV methods resulted in the following number of permutations:

- **Distance approach**: 18 (3× normalisation, 6× cost-matrix utilisation)
- **Qualitative approach**: 15 (3× normalisation, 5× cost-matrix utilisation)
- **Quantitative approach**: 3 (Only normalisation applicable)

The functional parameter permutations identified for each of these EGBV methods proved successful in providing depth to the concept of functional parameter sensitivity explored in this study.

7.2.2 Data compilation

The structure defined for the data compilation presented in Chapter 4 proved successful in providing an opportunity for repeated execution of the TEP model, which aided in the acquisition of multiple datasets to be used for the sensitivity analysis. The datasets acquired for the base case mode were only varied by the random seeds implemented for each condition simulated, and demonstrated sufficient variation to individual fault conditions simulated. Further preparation of the datasets acquired includes the transformation of process measurement data to attributed graphs containing the change in exergy and energy flow rates during system operation. The structure defined for data compilation additionally provided successful in presenting an opportunity to acquire datasets with the TEP model operation at Mode 5 to deliver near 30/70 product ratios. These were used in performing the validation of functional parameter sensitivity in Chapter 6. These datasets presented ample variation to fault conditions experienced, where a significant difference to respective operation conditions was observed by the EGBV methods.

7.2.3 Sensitivity analysis of EGBV functional parameters

The FDI executed to measure performance obtained from the EGBV method permutations delivered useful information, as the function parameter options demonstrated a significant deviation in the respective default permutations for these EGBV methods. It is worth noting the difference in MCC values delivered by the IQR-HEOM Threshold-based permutations for the Distance approach in comparison this method's default permutation. Additionally, various permutations delivered an improvement to their performance by their respective EGBV method's default permutation. The opposite was also observed, where some EGBV method permutations delivered considerably lower performance than their respective EGBV method's default permutation.

The sensitivity analysis performed using the FDI results for the base case mode operation of the model was able to identify significant sensitivity demonstrated towards the implemented functional parameters. The Distance approach also demonstrated significant sensitivity towards cost-matrix utilisation, while not being sensitive to normalisation implemented for each of the performance measures implemented. The Qualitative approach presents a varied case to sensitivity, where normalisation was identified as a significant source to variation for each of the performance measures, while sensitivity to cost-matrix utilisation was demonstrated in MCC and isolation values by this EGBV method. Lastly, the Quantitative approach only demonstrated sensitivity to normalisation in its MCC and accuracy results delivered.

7.2.4 Validation of sensitivity analysis

The validation performed through the use of an alternate operation mode for the TEP model delivered a significant contribution to both FDI results and sensitivity presented by each of the EGBV methods. The FDI performed using the Mode 5 datasets presented an opportunity to assess whether the performance demonstrated by permutations of the respective EGBV methods was consistent between operation modes, or if significant deviation could be observed. These observations were clear in the result comparisons for each performance measure individually for each of the EGBV methods, as discussed in Chapter 6.

The sensitivity analysis performed for validation purposes delivered useful results, where some sources of sensitivity demonstrated by the EGBV methods for the base case mode were validated as correct, and significant sensitivity was demonstrated to select sources by the EGBV methods that did not occur at base case mode operation. The sensitivity analysis performed as concept validation provided an additional layer of information to sensitivity demonstrated towards functional parameters, which would not have been available if only the base case mode had been implemented in this study.

7.3 Future research

With the validation performed in this study implementing the same model at an alternate operation mode, the concept of functional parameter sensitivity should be further explored by implementing a different system model, which would present a more complex attributed graph structure or a large quantity of fault conditions similar to the TEP model.

One permutation identified for the Distance approach demonstrated a significant improvement to this method's ability to consistently deliver a high MCC value, compared to the default permutation. It is recommended that the IQR-HEOM Threshold: 3-SD MIN permutation for the Distance approach is utilised in future research to determine whether this improved performance is valid when applied to a different system model.

7.4 Closure

The concept of functional parameter sensitivity demonstrated by EGBV methods was investigated in this study, and delivered valid results to conditions observed. FDI results delivered by the EGBV method permutations were analysed in respect to the functional parameters to establish whether significant sensitivity was demonstrated towards each. FDI performed by implementing the range of permutations for each of the EGBV methods demonstrated a deviation to performance delivered by the respective EGBV method's default permutations. The sensitivity identified towards functional parameters presented an indication on where focus should be placed in reducing the variation in performance delivered by EGBV methods in the future.

Bibliography

[1] T. Kurtoglu, O.J. Mengshoel, and S. Poll. A framework for systematic benchmarking of monitoring and diagnostic systems. In *2008 International Conference on Prognostics and Health Management*. IEEE, Oct 2008.

[2] M. Thirumarimurugan, N. Bagyalakshmi, and P. Paarkavi. Comparison of fault detection and isolation methods: A review. In *2016 10th International Conference on Intelligent Systems and Control (ISCO)*. IEEE, January 2016.

[3] P.M. Frank, E.A. Garcia, and B. Kppen-Seliger. Modelling for fault detection and isolation versus modelling for control. *Mathematical and Computer Modelling of Dynamical Systems*, 7(1):1–46, March 2001.

[4] K.R. Uren, G. van Schoor, and L. Auret. An energy-attributed graph approach for the purposes of FDI in a heated two-tank system. *IFAC - PapersOnLine*, 52(14):159–164, 2019.

[5] S. Greyling. *Graph-based fault detection for a gas-to-liquids process: an exergy approach.* PhD thesis, North-West University, 2020.

[6] J. Smith. Energy-based fault detection and isolation in an industrial steam turbine system. Master's thesis, North-West University, 2020.

[7] J.J. Downs and E.F. Vogel. A plant-wide industrial process control problem. *Computers and Chemical Engineering*, 17(3):245–255, Mar 1993.

[8] R.D. Braatz, L.H. Chiang, and E.L. Russell. *Fault Detection and Diagnosis in Industrial Systems*. Springer London, 2000.

[9] N.L. Ricker. Optimal steady-state operation of the Tennessee Eastman challenge process. *Computers and Chemical Engineering*, 19(9):949–959, 1995.

[10] T.J. McAvoy and N. Ye. Base control for the Tennessee Eastman problem. *Computers and Chemical Engineering*, 18(5):383–413, May 1994.

[11] P.R. Lyman and C. Georgakis. Plant-wide control of the Tennessee Eastman problem. *Computers and Chemical Engineering*, 19(3):321–331, Mar 1995.

[12] J. Vosloo, K.R. Uren, G. van Schoor, L. Auret, and H. Marais. Exergy-based fault detection on the Tennessee Eastman process. *IFAC - PapersOnLine*, 53(2):13713–13720, 2020.

[13] R. Isermann and P. Ballé. Trends in the application of model-based fault detection and diagnosis of technical processes. *Control Engineering Practice*, 5(5):709–719, May 1997.

[14] D. Miljkovic. Fault detection methods: A literature survey. In *MIPRO, 2011 Proceedings of the 34th International Convention, Opatija, Croatia, 23-27 May, 2011*, pages 750–755. IEEE, 2011.

[15] S. Jouili and S. Tabbone. Attributed graph matching using local descriptions. In J. Blanc-Talon, D. Popescu, W. Philips, and P. Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems - Acivs 2009*, Lecture Notes in Computer Science, Bordeaux, France, September 2009. SEE, Springer.

[16] S. van Graan, G. van Schoor, and K.R. Uren. Graph matching as a means to energy-visualisation of a counter-flow heat exchanger for the purpose of fault diagnosis. *IFAC-PapersOnLine*, 50(1):2842–2847, 2017. 20th IFAC World Congress.

[17] T.A. Reddy. Formulation of a generic methodology for assessing FDD methods and its specific adoption to large chillers. In *ASHRAE Transactions*, volume 113, Part 2, 2007.

[18] V. Venkatasubramanian, R. Rengaswamy, and S.N. Kavuri. A review of process fault detection and diagnosis part ii: Qualitative models and search strategies. *Computers and Chemical Engineering*, 27(3):313–326, Mar 2003.

[19] S. Frank, G. Lin, X. Jin, R. Singla, A. Farthing, and J. Granderson. A performance evaluation framework for building fault detection and diagnosis algorithms. *Energy and Buildings*, 192:84–92, Jun 2019.

[20] Stephen M. Frank, Guanjing Lin, Xin Jin, Rupam Singla, Amanda Farthing, Liang Zhang, and Jessica Granderson. Metrics and methods to assess building fault detection and diagnosis tools. Technical report, osti.gov, 3 2019.

[21] S. Razavi, A. Jakeman, A. Saltelli, C. Prieur, B. Iooss, E. Borgonovo, E. Plischke, S. Lo Piano, T. Iwanaga, W. Becker, S. Tarantola, J.H.A. Guillaume, J. Jakeman, H. Gupta, N. Melillo, G. Rabitti, V. Chabridon, Q. Duan, X. Sun, S. Smith, R. Sheikholeslami, N. Hosseini, M. Asadzadeh, A. Puy, S. Kucherenko, and H. R. Maier. The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling and Software*, 137:104954, Mar 2021.

[22] J.P.C. Kleijnen. Sensitivity analysis and optimization in simulation: Design of experiments and case studies. *Winter Simulation Conference Proceedings, 1995*, pages 133–140, 1995.

[23] A. Fassò and P.F. Perri. Sensitivity analysis. *Encyclopedia of Environmetrics*, 4:1968–1982, Jan 2002.

[24] E. Castillo, A.S Hadi, A. Conejo, and A. Fernández-Canteli. A general method for local sensitivity analysis with application to regression models and other optimization problems. *Technometrics*, 46(4):430–444, Nov 2004.

[25] A. Saltelli and P. Annoni. How to avoid a perfunctory sensitivity analysis. *Environmental Modelling and Software*, 25(12):1508–1517, Dec 2010.

[26] R.G. Sargent. An introductory tutorial on verification and validation of simulation models. In *2015 Winter Simulation Conference (WSC)*, pages 1729–1740, 2015.

[27] D.R. Wilson and T.R. Martinez. Improved heterogeneous distance functions. *J. Artif. Int. Res.*, 6(1):1–34, Jan 1997.

[28] M.S. Spencer, S.C. Bates Prins, and M.S. Beckom. Heterogeneous distance measures and nearest-neighbor classification in an ecological setting. *Missouri Journal of Mathematical Sciences*, 22(2), May 2010.

[29] P.I. Dalatu and H. Midi. Modified statistical approach for data preprocessing to improve heterogeneous distance functions. *Malaysian Journal of Mathematical Sciences*, 14(2):249–271, 2020.

[30] K. Iskandar, Noprianto, B.S. Abbas, B. Soewito, and R. Kosala. Two-Way ANOVA with interaction approach to compare content creation speed performance in knowledge management system. In *2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS)*, pages 1–5, 2016.

[31] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S.N. Kavuri. A review of process fault detection and diagnosis part i: Quantitative model-based methods. *Computers and Chemical Engineering*, 27(3):293–311, Mar 2003.

[32] D.C Lay. *Linear Algebra and Its Applications*. Addison-Wesley, fourth edition, 2012.

[33] H. Neser. *Energy-based visualisation of a Brayton cycle power conversion unit for the purpose of condition monitoring*. PhD thesis, North-West University, 2019.

[34] International Federation of Automation Control. Terminology in the Area of Fault Management.

[35] Qiuming Zhu. On the performance of Matthews Correlation Coefficient (MCC) for imbalanced dataset. *Pattern Recognition Letters*, 136:71–80, 2020.

[36] R.G. Shaw and T. Mitchell-Olds. Anova for unbalanced data: An overview. *Ecology*, 74(6):1638–1645, 1993.

[37] V. Ginot, S. Gaba, R. Beaudouin, F.A., and H. Monod. Combined use of local and anova-based global sensitivity analyses for the investigation of a stochastic dynamic model: Application to the case study of an individual-based model of a fish population. *Ecological Modelling*, 193(3):479–491, 2006.

[38] W. Wolmarans. A comparison of PCA- and energy-based FDI in a physical heated two-tank process. Master's thesis, North-West University, 2021.

Appendix A

Eigendecomposition - Standard deviation

Chapter 3 presents discussion on utilising the standard deviation of the cost-matrix as functional parameter for the Qualitative approach. The threshold by which qualitative evaluation would be performed in the Qualitative approach consists of calculating the standard deviation of the self-cost-matrix entries based for each respective reference graph. Similar to (Fig.2, [5]), (A.1) illustrates how the SD for each reference graph's self-cost matrix was calculated in this study.

$$\sigma = \sqrt{\frac{\Sigma(\lambda_{(REF)}(v) - \mu_\lambda)^2}{m}} \tag{A.1}$$

where $\lambda_{(REF)}(v)$ is each of the self-cost matrix values with $v = 1, ..., 15$ eigenvalues, and μ_λ the calculated eigenvalue average for the respective reference graph. The quantity of eigenvalues n is used as denominator before calculating the square root of the sum.

This delivers a standard deviation value σ individually for each of the reference graphs in an attempt to allow for a more discernible threshold between operation conditions.

Appendix B

Random variation seeds

This section presents the random variation seeds used during TEP model execution to acquire datasets for this study. From the dataset preparation discussed in chapter 4, 10 datasets were acquired for both base case mode- and mode 5 operation of the TEP model. These base case mode datasets were then utilised in chapter 5 in performing FDI and sensitivity analysis of the EGBV methods. Mode 5 datasets were then used in performing the validation seen in chapter 6 to confirm the initial results observed in chapter 5. These random variation seeds correspond between dataset numbers for both operation modes to ensure similarity of individual faults introduced during model operation.

Table B.1: Random seed values - Dataset 1

Condition	**Operation seed**	**Reference seed**
NOC	9902234324	8322308324
Fault 1	7984782901	7327843434
Fault 2	5784921734	5658678765
Fault 3	1723234455	4342232344
Fault 4	4346024432	4598956239
Fault 5	3433249064	5454589923
Fault 6	9445382439	9090909232
Fault 7	6788343442	6284545932
Fault 8	7654534567	9338398429
Fault 9	2589274931	2232323236
Fault 10	4356565463	6923255678
Fault 11	1940284333	1134345551
Fault 12	5683213434	5849489384
Fault 13	8934302332	8943242344
Fault 14	9873223412	9343430004
Fault 15	8943243993	5635346588
Fault 16	4243534565	1254545354
Fault 17	3485834345	3454354353
Fault 18	8998485332	8493323434
Fault 19	6678322168	6598593453
Fault 20	5457789234	1997072199

Table B.2: Random seed values - Dataset 2

Condition	**Operation seed**	**Reference seed**
NOC	4346024432	4598956239
Fault 1	5457789234	1997072199
Fault 2	8934302332	8943242344
Fault 3	8943243993	5635346588
Fault 4	7984782901	7327843434
Fault 5	2144342545	2132432423
Fault 6	5784921734	5658678765
Fault 7	1723234455	4342232344
Fault 8	2589274931	2232323236
Fault 9	4243534565	1254545354
Fault 10	4593493842	4545445883
Fault 11	6678322168	6598593453
Fault 12	1089278833	1039839281
Fault 13	4356565463	6923255678
Fault 14	1731738903	3420494299
Fault 15	3433249064	5454589923
Fault 16	7854912354	2994833239
Fault 17	3456432354	2891123453
Fault 18	6788343442	6284545932
Fault 19	7654534567	9338398429
Fault 20	3485834345	3454354353

Table B.3: Random seed values - Dataset 3

Condition	**Operation seed**	**Reference seed**
NOC	2144342545	2132432423
Fault 1	3485834345	3454354353
Fault 2	4593493842	4545445883
Fault 3	5683213434	5849489384
Fault 4	9902234324	8322308324
Fault 5	5784921734	5658678765
Fault 6	8943243993	5635346588
Fault 7	6678322168	6598593453
Fault 8	1940284333	1134345551
Fault 9	7984782901	7327843434
Fault 10	9445382439	9090909232
Fault 11	1723234455	4342232344
Fault 12	3456432354	2891123453
Fault 13	2589274931	2232323236
Fault 14	4356565463	6923255678
Fault 15	5457789234	1997072199
Fault 16	4346024432	4598956239
Fault 17	6788343442	6284545932
Fault 18	7654534567	9338398429
Fault 19	8934302332	8943242344
Fault 20	1089278833	1039839281

Table B.4: Random seed values - Dataset 4

Condition	**Operation seed**	**Reference seed**
NOC	1731738903	3420494299
Fault 1	9445382439	9090909232
Fault 2	1089278833	1039839281
Fault 3	1723234455	4342232344
Fault 4	8934302332	8943242344
Fault 5	1940284333	1134345551
Fault 6	7984782901	7327843434
Fault 7	3485834345	3454354353
Fault 8	5457789234	1997072199
Fault 9	6788343442	6284545932
Fault 10	5784921734	5658678765
Fault 11	9902234324	8322308324
Fault 12	2144342545	2132432423
Fault 13	8943243993	5635346588
Fault 14	3433249064	5454589923
Fault 15	5683213434	5849489384
Fault 16	7654534567	9338398429
Fault 17	9873223412	9343430004
Fault 18	4243534565	1254545354
Fault 19	4593493842	4545445883
Fault 20	4346024432	4598956239

Table B.5: Random seed values - Dataset 5

Condition	**Operation seed**	**Reference seed**
NOC	3456432354	2891123453
Fault 1	1089278833	1039839281
Fault 2	9902234324	8322308324
Fault 3	3485834345	3454354353
Fault 4	1731738903	3420494299
Fault 5	6788343442	6284545932
Fault 6	8943243993	5635346588
Fault 7	9445382439	9090909232
Fault 8	5683213434	5849489384
Fault 9	1723234455	4342232344
Fault 10	7984782901	7327843434
Fault 11	3433249064	5454589923
Fault 12	2589274931	2232323236
Fault 13	5784921734	5658678765
Fault 14	4593493842	4545445883
Fault 15	7654534567	9338398429
Fault 16	5457789234	1997072199
Fault 17	4356565463	6923255678
Fault 18	8934302332	8943242344
Fault 19	7854912354	2994833239
Fault 20	1940284333	1134345551

Table B.6: Random seed values - Dataset 6

Condition	Operation seed	Reference seed
NOC	8943243993	5635346588
Fault 1	3433249064	5454589923
Fault 2	2589274931	2232323236
Fault 3	3485834345	3454354353
Fault 4	7854912354	2994833239
Fault 5	4593493842	4545445883
Fault 6	5683213434	5849489384
Fault 7	1089278833	1039839281
Fault 8	6678322168	6598593453
Fault 9	5784921734	5658678765
Fault 10	9873223412	9343430004
Fault 11	4243534565	1254545354
Fault 12	3456432354	2891123453
Fault 13	7984782901	7327843434
Fault 14	1723234455	4342232344
Fault 15	6788343442	6284545932
Fault 16	8934302332	8943242344
Fault 17	9902234324	8322308324
Fault 18	2144342545	2132432423
Fault 19	1731738903	3420494299
Fault 20	5457789234	1997072199

Table B.7: Random seed values - Dataset 7

Condition	**Operation seed**	**Reference seed**
NOC	2589274931	2232323236
Fault 1	5784921734	5658678765
Fault 2	8943243993	5635346588
Fault 3	4356565463	6923255678
Fault 4	4346024432	4598956239
Fault 5	9445382439	9090909232
Fault 6	3456432354	2891123453
Fault 7	8934302332	8943242344
Fault 8	7984782901	7327843434
Fault 9	2144342545	2132432423
Fault 10	9873223412	9343430004
Fault 11	6788343442	6284545932
Fault 12	1731738903	3420494299
Fault 13	1723234455	4342232344
Fault 14	1089278833	1039839281
Fault 15	7854912354	2994833239
Fault 16	7654534567	9338398429
Fault 17	4593493842	4545445883
Fault 18	1940284333	1134345551
Fault 19	9902234324	8322308324
Fault 20	5457789234	1997072199

Table B.8: Random seed values - Dataset 8

Condition	**Operation seed**	**Reference seed**
NOC	9873223412	9343430004
Fault 1	6678322168	6598593453
Fault 2	8943243993	5635346588
Fault 3	7654534567	9338398429
Fault 4	2589274931	2232323236
Fault 5	7984782901	7327843434
Fault 6	3433249064	5454589923
Fault 7	3485834345	3454354353
Fault 8	5457789234	1997072199
Fault 9	7854912354	2994833239
Fault 10	8998485332	8493323434
Fault 11	1940284333	1134345551
Fault 12	8934302332	8943242344
Fault 13	3456432354	2891123453
Fault 14	1731738903	3420494299
Fault 15	6788343442	6284545932
Fault 16	1089278833	1039839281
Fault 17	2144342545	2132432423
Fault 18	4356565463	6923255678
Fault 19	5683213434	5849489384
Fault 20	9902234324	8322308324

Table B.9: Random seed values - Dataset 9

Condition	**Operation seed**	**Reference seed**
NOC	5784921734	5658678765
Fault 1	1723234455	4342232344
Fault 2	1731738903	3420494299
Fault 3	5683213434	5849489384
Fault 4	7854912354	2994833239
Fault 5	3433249064	5454589923
Fault 6	2144342545	2132432423
Fault 7	8934302332	8943242344
Fault 8	1940284333	1134345551
Fault 9	5457789234	1997072199
Fault 10	7984782901	7327843434
Fault 11	3456432354	2891123453
Fault 12	4593493842	4545445883
Fault 13	6788343442	6284545932
Fault 14	9873223412	9343430004
Fault 15	3485834345	3454354353
Fault 16	4346024432	4598956239
Fault 17	1089278833	1039839281
Fault 18	8943243993	5635346588
Fault 19	7654534567	9338398429
Fault 20	9902234324	8322308324

Table B.10: Random seed values - Dataset 10

Condition	**Operation seed**	**Reference seed**
NOC	4346024432	4598956239
Fault 1	5784921734	5658678765
Fault 2	6788343442	6284545932
Fault 3	2144342545	2132432423
Fault 4	3433249064	5454589923
Fault 5	2589274931	2232323236
Fault 6	3456432354	2891123453
Fault 7	8943243993	5635346588
Fault 8	7854912354	2994833239
Fault 9	5683213434	5849489384
Fault 10	8998485332	8493323434
Fault 11	8934302332	8943242344
Fault 12	4356565463	6923255678
Fault 13	4593493842	4545445883
Fault 14	7984782901	7327843434
Fault 15	9445382439	9090909232
Fault 16	6678322168	6598593453
Fault 17	3485834345	3454354353
Fault 18	1089278833	1039839281
Fault 19	9873223412	9343430004
Fault 20	9902234324	8322308324

Appendix C

TEP model datasets

The TEP model datasets acquired in this study include 10 datasets simulated at base case mode operation, and 10 datasets simulated at mode 5 operation of the model. Chapters 4 and 6 include discussion on the structure of these datasets where the duration values utilised during simulation of each mode is included below:

Datasets - Base case mode

- Duration of simulated reference condition. - 25 Hours
- Duration of simulated operation condition. - 48 Hours
- Duration of normal operation prior to reference condition start. - 1 Hour
- Duration of normal operation prior to operation condition start. - 8 Hours

Datasets - Mode 5

- Duration of simulated reference condition. - 25 Hours
- Duration of simulated operation condition. - 48 Hours
- Duration of normal operation prior to reference condition start. - 8 Hours
- Duration of normal operation prior to operation condition start. - 8 Hours

All TEP model datasets generated in this study are freely accessible at the links below:

- **Base case mode:** **https://figshare.com/s/de29bc671da201ce8e30**
- **Mode 5:** **https://figshare.com/s/493088a37a1c07291f1d**

Appendix D

MATLAB code - EGBV methods

The MATLAB code developed for this study consist of multiple files for the respective EGBV methods and their associated permutations. Each EGBV method was separated into 3 files based on their normalisation values where each of these files include execution of the associated cost-matrix utilisation functional parameter values discussed in chapter 3. The results demonstrated in chapters 5 and 6 were obtained from the calculations executed in these files.

All MATLAB code files developed in this study are freely accessible at the link below:

- **MATLAB Code:** https://figshare.com/s/fd09de01727afe94d921