

**DEVELOPMENT OF GAS DENSITY MEASUREMENT MODEL (TANK)
BASED ON PRINCIPLE COMPONENT ANALYSIS (PCA) TECHNIQUE**

WAN MOHD HIDAYAT BIN WAN ABDULLAH

KOLEJ UNIVERSITI KEJURUTERAAN DAN TEKNOLOGI MALAYSIA

**KOLEJ UNIVERSITI KEJURUTERAAN DAN
TEKNOLOGI MALAYSIA**

BORANG PENGESAHAN STATUS TESIS

JUDUL: **DEVELOPMENT OF GAS DENSITY MEASUREMENT MODEL (TANK)
BASED ON PRINCIPAL COMPONENT ANALYSIS (PCA) TECHNIQUE**

SESI PENGAJIAN: **2005/2006**

Saya **WAN MOHD HIDAYAT BIN WAN ABDULLAH**
(HURUF BESAR)

mengaku membenarkan tesis (PSM/Sarjana/Doktor Falsafah)* ini disimpan di Perpustakaan Kolej
Universiti Kejuruteraan dan Teknologi Malaysia dengan syarat-syarat kegunaan seperti berikut:

1. Tesis adalah hakmilik Kolej Universiti Kejuruteraan dan Teknologi Malaysia.
2. Perpustakaan Kolej Universiti Kejuruteraan dan Teknologi Malaysia dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. **Sila tandakan (✓)

☐

SULIT

(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

☐

TERHAD

(Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

☒

TIDAK TERHAD

Disahkan oleh

(TANDATANGAN PENULIS)

Alamat Tetap: **FPSB, LEPAR UTARA 07**
27000 JERANTUT
PAHANG DARUL MAKMUR

(TANDATANGAN PENYELIA)

MOHD YUSRI B. MOHD YUNUS
(Nama Penyelia)

Tarikh: **20 November 2006**

Tarikh: **20 November 2006**

CATATAN:

* Potong yang tidak berkenaan.

** Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh tesis ini perlu dikelaskan sebagai SULIT atau TERHAD.

◆ Tesis dimaksudkan sebagai tesis bagi Ijazah Doktor Falsafah dan Sarjana secara penyelidikan, atau disertai bagi pengajian secara kerja kursus dan penyelidikan, atau Laporan Projek Sarjans Muda (PSM).

“I hereby acknowledge that I had read this technical writing and in my opinion this technical writing is sufficient in terms of scope and quality for the purpose of the granting of Bachelor of Chemical Engineering.”

Signature	:
Name of Supervisor	: Tuan Haji Mohd Yusri bin Mohd Yunus
Date	: 22 nd November 2006

DEVELOPMENT OF GAS DENSITY MEASUREMENT MODEL (TANK) BASED
ON PRINCIPAL COMPONENT ANALYSIS (PCA) TECHNIQUE

WAN MOHD HIDAYAT BIN WAN ABDULLAH

A thesis submitted in fulfillment
of the requirements for the award of the degree of
Bachelor of Chemical Engineering

Faculty of Chemical & Natural Resources Engineering
University College of Engineering & Technology Malaysia

November 2006

DECLARATION

I declare that this thesis entitled “*Development Gas Density Measurement Model (Tank) Based on Principal Component Analysis (PCA) Techniques*” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :
Name : WAN MOHD HIDAYAT BIN WAN ABDULLAH
Date : 22nd NOVEMBER 2006

*To my beloved mother and father...
Thank you for bringing me onto this beautiful world
My love...you are always in my mind*

ACKNOWLEDGEMENT

As a vicegerent who is being created, I put myself down to express my highest thankful to Allah S.W.T for giving me strength and spirit to complete this final project with full of pride and dignity.

I would like to express my deepest gratitude to the following persons for their unlimited and kindness help as well as guidance enabled me to complete this research project in time as a partial fulfillment of the requirement of the degree of Bachelor Engineering (Chemical).

To Tuan Hj. Mohd Yusri B. Mohd Yunus, my lovely supervisor, thanks a lot for helping me during the progress of the research project, either in a way of searching the required information as well as guiding me to be a good person. Without his generosity in sparing his precious time to guide me and answer my doubts, the aim of project may not be fulfilled.

My highest appreciation to my much closed friends Mimiey, Faiss, Deqwan, Rufizal, and Anuar who inspired me a lot and in addition I am very grateful to my family members, no amount of gratitude could repay their kindness of being there as well as the patience to put up with my whim.

I also would like to thank my fellow course mates and friends for their assistances, advices and ideas in producing a resourceful, fruitful and practicable research. May God bless all of you.

ABSTRACT

The real importance of controlling gas density is based on the fact that it contributes significantly to the composition of the gas being controlled. Even with small variations can cause significant process upsets if not adequately controlled due to changing in composition. This research purposely to develop a model of gas density measurement as the process instrument nowadays hardly to correlate simultaneously the input variables to the output variables. In this respect, density is known as output whereby it depends on other variables (input) which are pressure and temperature of the given process. In other words, all variables which can be directly measured and contribute to the primary variables are known as 'secondary variables'. Thus, the Principal Component Analysis (PCA) approach is used plus the MATLAB programming to ensure the gas density measurement model is valid as the real process. At the end of this research, the developed model can be used to identify the best configurations of inputs variables in producing the specified output that intended. The best outcome is the effect of temperature and pressure to the density is very significant as the relationships between those variables are proportional to each other.

ABSTRAK

Kepentingan utama dalam pengawalan ketumpatan gas adalah menyangkan secara signifikan komposisi gas yang dikawal. Walau dengan perbezaan yang kecil, ianya boleh menyebabkan proses tidak menepati seperti yang dikawal melihat kepada perubahan komposisi. Kajian ini bertujuan membentuk sebuah model pengiraan ketumpatan gas seperti kerana seperti yang kita ketahui, proses instrumen hari ini sukar untuk menghubungkan secara serentak pembolehubah masuk terhadap pembolehubah keluar dalam satu sesi proses yang sama. Dalam kaian ini, ketumpatan adalah pembolehubah keluar manakala suhu dan tekanan adalah pembolehubah masuk. Dalam erti kata yang lain, pembolehubah yang boleh dikira secara langsung merupakan pembolehubah sekunder. Oleh itu, pendekatan Analisis Prinsip Komponen (PCA) bersama simulasi MATLAB digunapakai untuk memastikan model pengiraan ketumpatan gas boleh memberikan bacaan seperti proses sebenar. Keperbagaian model yang sudah dimajukan atau diperkembangkan pada akhir projek ini boleh digunakan untuk mengenalpasti konfigurasi pembolehubah keluar seperti yang dikehendaki. Keputusan menunjukkan, pembolehubah keluar bergantung secara langsung dengan pembolehubah masuk kerana hubungan di antara pembolehubah masuk dan pembolehubah keluar adalah berkadaran secara langsung.

TABLES OF CONTENTS

CHAPTER	TITLE	PAGE
	TITLE PAGE	i
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGEMENT	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLES OF CONTENT	vii
	LIST OF TABLES	x
	LIST OF FIGURES	xi
	LIST OF ABBREVIATION OR NOMENCLATURE	xii
	LIST OF APPENDICES	xiii

I	INTRODUCTION	
1.0	Overview and Fundamental of Ideal Gas Law	1
1.0.1	Ideal Gas Law with Constraints	2
1.1	Problem Statement	4
1.2	Objectives	5
1.3	Scopes of Study	5
1.4	Research Contribution	5
1.5	Summary of Report	6
II	FUNDAMENTAL THEORY	
2.0	Introduction	7
2.1	Gas Density Measurement	9
2.1.1	Composition of Air	10
2.2	Principal Component Analysis (PCA)	11
2.2.1	Original Data Matrix, Variance-Covariance Matrix and Correlation Matrix	16
2.2.2	Determine the Eigenvalue and Eigenvector Of the Matrix	19
2.2.3	Number of Principal Component Required	20
II	LITERATURE RIVIEW	
3.0	Introduction	24
3.1	Gas Density Measurement	24
3.2	Principal Component Analysis (PCA)	29
IV	METHODOLOGY	
4.0	Introduction	35
4.1	Case Study of AFPT 921	35
4.2	Methodology Framework	37

	4.2.1 Data Gathering	38
	4.2.2 Density Calculation	38
	4.2.3 Plot Profile	38
	4.2.4 Model Development	38
	4.2.5 Validation	39
V	RESULT AND DISCUSSION	
	5.0 Data Gathering	40
	5.1 Gas Density Measurement	43
	5.2 Plot Profile	46
	5.3 Development of Principal Component Analysis Model	47
	5.4 The PCA Codes	49
	5.5 Validation	51
	5.6 Analysis and Discussion	54
VI	CONCLUSION	
	6.0 Conclusions	55
	6.1 Recommendation	55
	REFERENCES	56
	APPENDIX I	59
	APPENDIX II	63

LIST OF TABLES

TABLES NO.	TITLE	PAGE
2.1	The Composition of Dry Air is Relatively Uniform Around the World.	10
5.1	Temperatures in Pipeline (Experiment Data)	37
5.2	Pressures in Pipeline (Experiment Data)	42
5.3	Results of Density Calculation	45
5.4	Results of Gas Density from Programming	51
5.5	Comparison between Manual and Simulation	53

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
1.1	PvT Surface	3
2.1	PCA Graphical matrix representation	13
2.2	System for Two Variables Distribution	15
2.3	Graphical Representation for PCA	15
2.4	Scree Plot	22
2.5	Principal Component Analysis Procedures	23
4.1	Flowcharts for Methodology	37
5.1	Profile Density versus Temperature	46
5.2	Profile Density versus Pressure	47
5.3	The Programming Data Set in M-file	48
5.4	The Programming Data Set in M-file (contd)	48

LIST OF ABBREVIATION OR NOMENCLATURE

T	temperature in pipeline, °C
P	pressure in pipeline, Pa
ρ	density in tank, kg/m ³
X	data matrix, consisting of the set of all data vectors, one vector per column
N	the number of column vectors in the data set
M	the number of elements in each column vector
μ	vector of empirical means, one mean for each row m of the data matrix
s	vector of empirical standard deviations, one standard deviation for each row m of the data matrix
Z	z-scores, computed using the mean and standard deviation for each row m of the data matrix
C	covariance matrix
R	correlation matrix
V	matrix consisting of the set of all eigenvectors of C , one eigenvector per Column
W	matrix of basis vectors, one vector per column, where each basis vector is one of the eigenvectors of C , and where the vectors in W are a sub-set of those in V
Y	matrix consisting of N column vectors, where each vector is the projection of the corresponding data vector from matrix X onto the basis vectors contained in the columns of matrix W .
Cov	Covariance
Eig	Eigenvalue

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
I	CASE STUDE OF AIR FLOW PRESSURE TEMPERATURE (AFPT921 PLANT)	59
II	CODING FOR GAS DENSITY MEASUREMENT MODEL USING PCA TECHNIQUE IN MATLAB	63

CHAPTER I

INTRODUCTION

1.0 Overview and Fundamental of Ideal Gas Law

An ideal gas is defined as one in which all collisions between atoms or molecules are perfectly elastic and in which there are no intermolecular attractive forces. One can visualize it as a collection of perfectly hard spheres which collide but which otherwise do not interact with each other. In such a gas, all the internal energy is in the form of kinetic energy and any change in internal energy is accompanied by a change in temperature.

An ideal gas can be characterized by three state variables: absolute pressure (P), volume (V), and absolute temperature (T). The relationship between them may be deduced from kinetic theory and is called the Ideal Gas Law

$$PV = nRT = NKT \quad (1.1)$$

Where,

n = number of moles

R = universal gas constant = 8.3145 J/mol K

N = number of molecules

k = Boltzmann constant = 1.38066×10^{-23} J/K = 8.617385×10^{-5} eV/K

$k = R/N_A$

N_A = Avogadro's number = 6.0221×10^{23} /mol

The ideal gas law can be viewed as arising from the kinetic pressure of gas molecules colliding with the walls of a container in accordance with Newton's laws.

But there is also a statistical element in the determination of the average kinetic energy of those molecules. The temperature is taken to be proportional to this average kinetic energy; this invokes the idea of kinetic temperature. One mole of an ideal gas at STP occupies 22.4 liters.

1.0.1 Ideal Gas Law With Constraints

For the purpose of calculations, it is convenient to place the ideal gas law in the form:

$$\frac{P_i V_i}{T_i} = \frac{P_f V_f}{T_f} \quad (1.2)$$

where the subscripts i and f refer to the initial and final states of some process. If the temperature is constrained to be constant, this becomes:

$$P_i V_i = P_f V_f \quad (1.3)$$

which is referred to as Boyle's Law. If the pressure is constant, then the ideal gas law takes the form:

$$\frac{V_i}{T_i} = \frac{V_f}{T_f} \quad \text{or} \quad V_f \frac{T_f}{T_i} \quad (1.4)$$

which has been historically called Charles' Law. It is appropriate for experiments performed in the presence of a constant atmospheric pressure.

All the possible states of an ideal gas can be represented by a PvT surface as illustrated Figure 1.1. The behavior when any one of the three state variables is held constant is also shown.

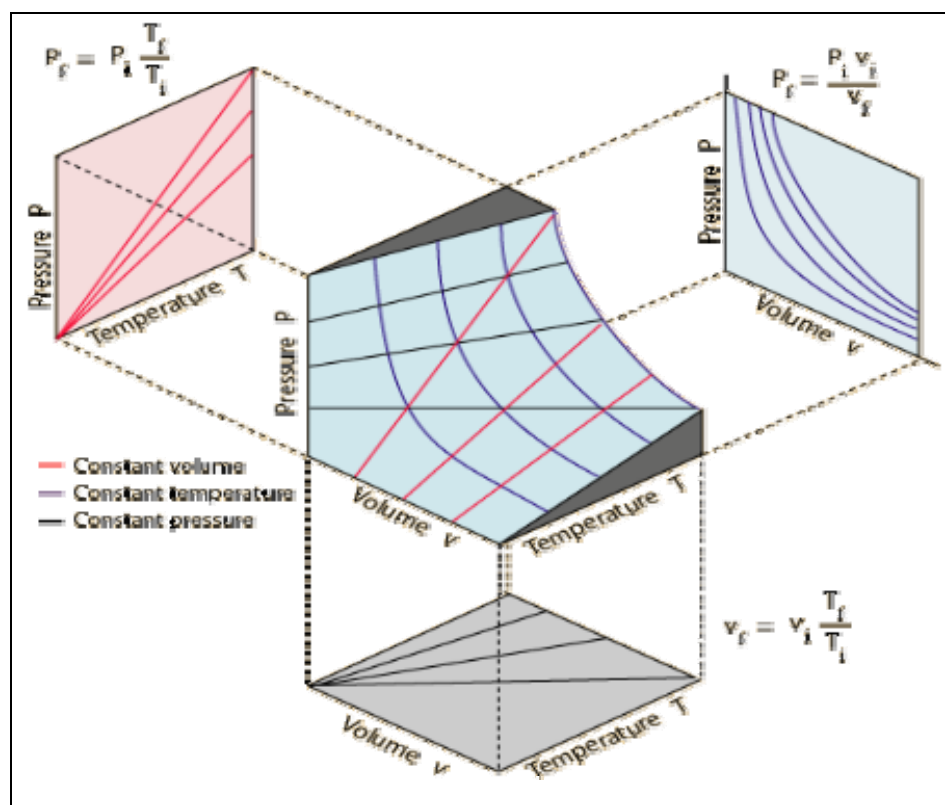


Figure 1.1 PvT Surface

1.1 Problem Statement

The real importance of controlling gas density is based on the fact that it contributes significantly to the composition of the gas being controlled. Even with small variations can cause significant process upsets if not adequately controlled due to changing in composition. As for example, the variations in gas composition experienced resulted in the production engineers using a higher oxygen level than was actually needed to allow safety margin. As a result, this will lead to lower process efficiency and higher production cost [1].

However, density cannot be measured directly from the process due to several reasons as the followings [2].

- a) Lack of appropriate on-line instrumentation.
 - i) Process operation has to depend on laboratory analysis.
 - ii) The cost of installing a sensor may not be justified.
- b) Reliability of on-line instruments.
 - i) Slow feedback.
 - ii) Lack of measurement in a timely manner.

In this respect, density is known as 'primary variables' whereby it depends on other variables (secondary variables) which are pressure and temperature of the given process. In other words, all variables which can be directly measured and contribute to the primary variables are known as 'secondary variables'. As a result, the quality of the product cannot be measured and analyzed in on-line or real time manner and this could lead to off-specification products.

Besides that, those primary variables are also not being monitored and controlled thoroughly as well as simultaneously with all the main secondary variables that available in the plant. Eventually, the analysis being carried out is lacking of reliability aspect, in the sense that it does not incorporate all secondary variables that potentially affect upon the primary variables when changes occur.

1.2 Objectives

- i. The aims of this study is to model the relationships between quality output (density) and secondary outputs (temperature and pressure) using Principle Component Analysis (PCA).
- ii. To develop a program to measure the density from AFPT 921 plant.

1.3 Scopes of Study

There are five major scopes that have been identified in order to achieve the objectives of this research:

- i) Data gathering from Air-Flow-Pressure-Temperature (AFPT) Plant to obtain a set of secondary outputs data.
- ii) Calculate the primary output by using Microsoft Excel.
- iii) To develop variable profile
- iv) Modeling the gas density measurement by using PCA.
- v) Comparing performance result of the gas density measurement modeled with the conventional measurement available.

1.4 Research Contribution

The main outputs of this research are:

- i) A new gas density measurement model is proposed and developed by using PCA.
- ii) A new programming on gas density measurement is developed.

1.5 Summary of Report

This thesis contains five chapters. The first chapter comprises of the introduction of the research, fundamental of Ideal Gas Law, objectives of the research, scope of the research and contributions of this research. Chapter two reviews the fundamental theory of gas density measurement and Principal Component Analysis (PCA).

Chapter three consists of the literature review which is reveals of the previous research and study about gas density measurement and the Principal Component Analysis (PCA)

Chapter four expresses of the research methodology which is in this chapter, the case study on AFPT 921 plant has been made while method algorithm is developed as a guidance to complete the research.

Chapter five presents the result and discussion of the research. The results are systematically presented into the tables and graph togetherness with the discussion and justification.

CHAPTER II

FUNDAMENTAL THEORY

2.0 Introduction

Model AFPT 921 is a process control training system that uses only air to simulate gas, vapor or steam. Air is readily available from a compressor. It provides the simple gas physical processes where the measurement and control of their important variables of flow, temperature and pressure can be studied.

In general, density is one of gas physical properties in which it is very much dependant on the pressure and temperature of the given process under consideration. This is very true as illustrated in the Perfect Gas Law as follows:

$$d = \frac{m}{V} = \frac{P}{T} \cdot \frac{M}{R} = \frac{P}{T} \times \text{constant} \quad (2.1)$$

Whereby,

P = Pressure

V = Volume

m = Mass

M = Molecular weight

R = 8.314 J/mol.K

T = Temperature

$$\text{Therefore, } d = \frac{m}{V} = \frac{P}{T} \cdot \frac{M}{R} = \frac{P}{T} \times \text{constant} \quad (2.2)$$

for any given gas. For instance, the following shows the real issue in measuring gas density.

Assume the total quantity of gas through a pipeline is held constant at 50 kg/hr but the flowing pressure and temperature can be varying. Initially, the flow rate is at the Normal Condition (NTP) of 0°C temperature and 1 atm pressure, and increase gradually to 120°C and 1.485 atm respectively. As a result, the actual volume of the gas has expanded as well as contracted due to temperature and pressure changes at the same time.

Whilst the initial gas density is 1.293 kg/Nm³, the new calculated density could be calculated approximately as follows:

Initial density = 1.293 kg/Nm³ at 0°C (273°K) and 1 atm (NTP) .

$$\text{Therefore, new density} = 1.293 \times \frac{273\text{K}}{393\text{K}} \times \frac{1.485 \text{ atm}}{1.0 \text{ atm}}$$

$$= 1.334 \text{ kg/ m}^3 \text{ at } 120^\circ\text{C, } 1.485 \text{ atm}$$

2.1 Gas Density Measurement

The gas density can be measured by many ways. Nowadays there are a lot of equipment were invented for the purpose mean. But when it comes to the basic element which is reached to the assuming of ideal gas, it is definitely has to reveal the characteristics of the ideal gas.

The ideal gas law can be manipulated to find the density of a substance, the manipulation of the ideal gas law as shown below:

$$PV=nRT \quad n=\frac{m}{mw}, \quad \rho=\frac{m}{V}$$

Rearrange the equation,

$$PV=\frac{m}{mw}RT \quad (2.3)$$

$$\frac{m}{V}=\frac{mw}{R}\left(\frac{P}{T}\right)$$

$$\rho = \text{constant} \left(\frac{P}{T} \right)$$

In several cases, the P and T is to be manipulated as the density very dependant of changing the pressure and temperature.

2.1.1 Composition of Air

Air is a mixture of gases with overall physical and chemical properties consistent with its composition. The "real" air that we encounter on a day-to-day basis, and use for breathing, burning fuels for heating, transportation, power generation and many other purposes, can be thought of as a mixture of three types of components:

- i) Standard dry air, which is mainly composed of three gases. They are nitrogen (about 78%), oxygen (about 21%), and argon (a bit less than 1%). All three of these gases can be economically recovered as industrial gas products. Standard dry air also contains a small amount of carbon dioxide, and very small amounts of neon, helium, krypton, hydrogen and xenon

Table 2.1 The composition of dry air is relatively uniform around the world.

Standard Dry Air Composition (Detailed Analysis):					
Gas	% by Volume	% by Weight	Parts per Million (V)	Chemical Symbol	Molecular Weight
Nitrogen	78.08	75.47	780805	N ₂	28.01
Oxygen	20.95	23.20	209450	O ₂	32.00
Argon	0.93	1.28	9340	Ar	39.95
Carbon Dioxide	0.038	0.0590	380	CO ₂	44.01
Neon	0.0018	0.0012	18.21	Ne	20.18
Helium	0.0005	0.00007	5.24	He	4.00
Krypton	0.0001	0.0003	1.14	Kr	83.80
Hydrogen	0.00005	Negligible	0.50	H ₂	2.02
Xenon	8.7 x 10 ⁻⁶	0.00004	0.087	Xe	131.30

2.2 Principal Component Analysis

Principal Components Analysis (Jackson, 1980; Geladi and Kowalski, 1986; Wold *et al.*, 1987; Jackson, 1991; Johnson and Wichern 1992) can be described via using mathematical representation and graphical representation. From the mathematical point of view, Principal Components Analysis is a multivariate technique, which can transform a set of original variables x_1, x_2, \dots, x_m to a set of new variables P_1, P_2, \dots, P_m . Actually, the newly formed variables are called Principal Components (PC), which are the linear combinations of the original variables. The mathematical representations that describe the transformation are shown as follow:

$$P_1 = v_{1,1}x_1 + v_{1,2}x_2 + \dots + v_{1,m}x_m$$

$$P_2 = v_{2,1}x_1 + v_{2,2}x_2 + \dots + v_{2,m}x_m$$

.

.

$$P_m = v_{m,1}x_1 + v_{m,2}x_2 + \dots + v_{m,m}x_m \quad (2.4)$$

$$P_{m \times n} = V_{m \times m} X_{m \times n}^T \quad (2.5)$$

The original matrix, X has m variables with each variable has n measurements. The data are arranged in the form of $n \times m$, where the measurements of a variable are organized in the form of column vector. A row vector, which contains the measurement of all variables is denoted as $x_i = [x_{i,1} \ x_{i,2} \ \dots \ x_{i,m}]$, while $[x_{i,1} \ x_{i,2} \ \dots \ x_{i,m}]^T$ is the transpose of the row vector. V is known as the *eigenvector* matrix. V is the weighting matrix in forming the linear combinations of the original variables. The *eigenvectors* matrix, V , contains *eigenvectors* or loading vectors v_1, v_2, \dots, v_m . Each *eigenvector* V_j is a row vector which contains the arrangement of elements as $V_j = [v_{j,1} \ v_{j,2} \ \dots \ v_{j,m}]$. The *eigenvectors* matrix can be shown in the following matrix:

$$\text{Eigenvectors matrix, } \mathbf{V} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_m \end{bmatrix} = \begin{bmatrix} v_{1,1} & v_{1,2} & \dots & v_{1,m} \\ v_{2,1} & v_{2,2} & \dots & v_{2,m} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ v_{m,1} & v_{m,2} & \dots & v_{m,m} \end{bmatrix} \quad (2.6)$$

\mathbf{P} is the principal components scores matrix. A particular principal component score, $P_{i,j}$, where $j < m$, is obtained via the product of a row vector variable measurement $[x_{i,1} \ x_{i,2} \ \dots \ x_{i,m}]$ to the particular *eigenvector* $\mathbf{V_j}$, which is shown in the following equation:

$$P_{i,j} = \mathbf{v}_j \mathbf{x}_i^T \quad (2.7)$$

$$P_{i,j} = \begin{bmatrix} v_{j,1} & v_{j,2} & \dots & v_{j,m} \end{bmatrix} \cdot \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \cdot \\ \cdot \\ x_{i,m} \end{bmatrix} \quad (2.8)$$

$$\mathbf{P}_{j(1 \times n)} = \mathbf{v}_{j(1 \times m)} \mathbf{X}_{(m \times n)}^T \quad (2.9)$$

$\mathbf{p_j}$ is the score vector for the Principal Component j , while $\mathbf{v_j}$ is the *eigenvector* j in order to form Principal Component. Equation 2.8 is used to determine the principal component score matrix \mathbf{P} , which contains n scores for each principal component:

$$\mathbf{P}_{(m,n)} = \mathbf{V}_{(m \times m)} \mathbf{X}_{(m \times n)}^T \quad (2.10)$$

On the other hand, the original variables can also be represented in term of newly formed variables as follow and as shown in Figure 2.1:

$$\mathbf{X}_{(n \times m)} = \mathbf{P}_{1(n \times l)}^T \mathbf{V}_{1(l \times m)} + \mathbf{P}_{2(n \times l)}^T \mathbf{V}_{2(l \times m)} + \dots + \mathbf{P}_{m(n \times l)}^T \mathbf{V}_{m(l \times m)} \quad (2.11)$$

$$\mathbf{X}_{(n \times m)} = \mathbf{P}_{(n \times m)}^T \mathbf{V}_{(m \times m)} \quad (2.12)$$

$$\begin{aligned} \begin{bmatrix} \mathbf{X} \end{bmatrix}_n^m &= \begin{bmatrix} \mathbf{p}_1^T \end{bmatrix}_n^1 \begin{bmatrix} \mathbf{v}_1 \end{bmatrix}_1^m + \begin{bmatrix} \mathbf{p}_2^T \end{bmatrix}_n^1 \begin{bmatrix} \mathbf{v}_2 \end{bmatrix}_1^m + \dots + \begin{bmatrix} \mathbf{p}_m^T \end{bmatrix}_n^1 \begin{bmatrix} \mathbf{v}_m \end{bmatrix}_1^m \\ &= \begin{bmatrix} \mathbf{p}^T \end{bmatrix}_n^m \begin{bmatrix} \mathbf{v} \end{bmatrix}_m^m \end{aligned}$$

Figure 2.1 PCA graphical matrix representation

Where \mathbf{X}	: original data matrix	dimension: $n \times m$
$\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_m$: eigenvectors	dimension: $1 \times m$
\mathbf{V}	: eigenvector matrix	dimension: $m \times m$
\mathbf{V}_a	: matrix that contains a eigenvectors	dimension: $a \times m$
\mathbf{P}_j	: Principal Component j score vector	dimension: $1 \times n$
\mathbf{P}	: Principal Component score matrix	dimension: $m \times n$
\mathbf{P}_a	: Principal Component score matrix	dimension: $a \times n$
for PC I till PC a		

From Equation 2.1, the number of original variables, x_m equals to the number of the newly formed variables, P_m . If all Principal Components are used to represent the original variables, the original raw data matrix is reproduced as shown in Equation 2.9. If this is the condition, the purpose for using Principal Components Analysis as dimension reduction technique will be lost. In order to maintain the uniqueness of this

technique, only several Principal Components will be used to represent most of the original data variation. If a Principal Components are decided to retained with $a < m$, then Equation 2.9 can be written as follow:

$$\mathbf{X} = \mathbf{P}_a^T \mathbf{V}_a + \mathbf{P}_{m-a}^T \mathbf{V}_{m-a} \quad (2.13)$$

$$\mathbf{X} = \mathbf{P}_a^T \mathbf{V}_a + \mathbf{E} \quad (2.14)$$

The retained principal components $[P_1, P_2, \dots, P_a]$, which form the $P; V_a$ term, are associated with systematic variation in data while the residual principal components $[P_{a+1}, P_{a+2}, \dots, P_m]$, which form the residual matrix \mathbf{E} are considered of containing measurement errors (Seborg *et al.*, 1996). Therefore, PCA is a multivariate analysis technique that could use less number of newly formed variables to represent the original data variations without losing significant information. Information here is referred to data variation.

For the graphical representation of PCA, the linear combinations of the original variables in forming the new variables are actually representing selection of a new coordinate system with $[P_1, P_2, \dots, P_m]$ as the new axes obtained by rotating the original system with x_1, x_2, \dots, x_n as the coordinate axes. The new axes represent the direction with maximum variability and provide a simpler and more parsimonious description of the variance-covariance matrix or correlation matrix (Johnson and Wichern, 1992). Figure 2.2 and 2.3 are prepared to give graphical representations of PCA.

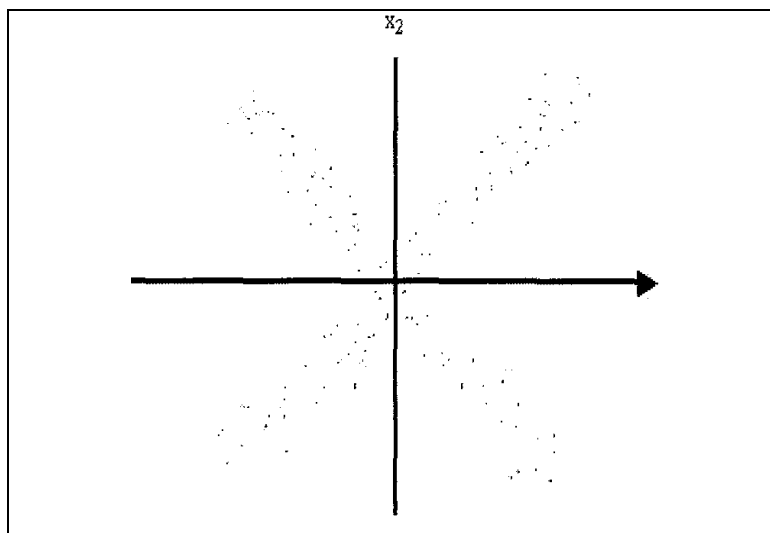


Figure 2.2 System for two variables distribution

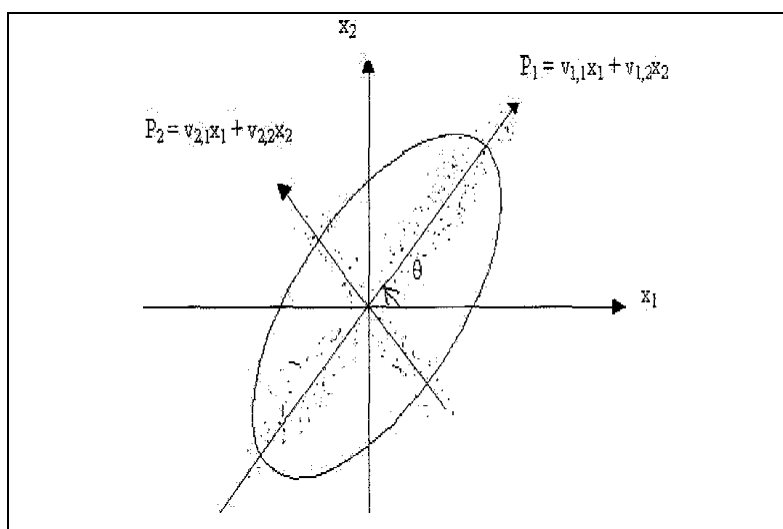


Figure 2.3 graphical representations for PCA

For example, Figure 2.2 shows a system has two variables x_1 and x_2 . In order to have a better representation of the system variability in term of linear combinations of x_1 and x_2 , the original axes are rotated to θ degree so that the newly formed Principal Component 1 will represent the maximum variation of the system. Maximum variation is achieved if the perpendicular distance of all the points to the line is minimum. The Principal Component 2 will represent the maximum variation of the remaining variation. The process of forming new variables is repeated until all the Principal Components are determined for the system, which has more than two variables. The representation of axes rotation to form Principal Components is presented in Figure 2.3. Jackson (1980) described that the principal axes (P_1 and P_2) are the rotation of the covariance matrix or correlation matrix, and the *eigenvectors* (coefficients of principal components) are the direction cosines of the new axes related to the old. The graphical representation for more than three axes is impossible.

2.2.1 Original Data Matrix, Variance-Covariance Matrix and Correlation Matrix

The initial step of Principal Components Analysis is to have a set of data and arrange in the original matrix \mathbf{X} , which has n measurements for m variables. Actually, PCA enable the variance-covariance structure or correlation structure of the original data to be explained by forming the several linear combinations of original variables because the *eigenvectors* that used to form the new variables are determined from the variance-covariance matrix or correlation matrix of original variables (Johnson and Wichern, 1992). Variance is a measure of a variable's dispersion or variation while covariance is a measure of co-variation between two variables. The variance and covariance of each variable are determined respectively from Equation 2.12 and 2.13. The correlation $r_{j,k}$ is a measure of linear association between two variables like variable j and variable k . The correlation between variables and correlation matrix are determined from Equation 2.15 and 2.16 accordingly.

$$\text{Variance, } s^2_{j,j} = \frac{\sum_i^n (x_{i,j} - \bar{x}_j)^2}{n-1} \quad (2.15)$$

$$\text{Covariance, } s_{j,k} = \frac{\sum_i^n (x_{i,j} - \bar{x}_j) \sum_k^n (x_{i,k} - \bar{x}_k)}{(n-1)(n-1)} \quad (2.16)$$

$$\text{Variance-covariance matrix, } S = \begin{bmatrix} s_{1,1} & s_{1,2} & \cdots & s_{1,m} \\ s_{2,1} & s_{2,2} & \cdots & s_{2,m} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ s_{m,1} & s_{m,2} & \cdots & s_{m,m} \end{bmatrix} \quad (2.17)$$

$$\text{Correlation, } r_{j,k} = \frac{s_{j,k}}{\sqrt{s_{j,j}} \sqrt{s_{k,k}}} \quad (2.18)$$

$$\text{Correlation matrix, } R = \begin{bmatrix} r_{1,1} & r_{1,2} & \dots & r_{1,m} \\ r_{2,1} & r_{2,2} & \dots & r_{2,m} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ r_{m,1} & r_{m,2} & \dots & r_{m,m} \end{bmatrix} \quad (2.19)$$

Where: i : i-th measurement $\overline{x_k}$: mean value for variable x_j

$s^2_{j,j}$: variance of variable x_j $s_{j,k}$: covariance of variable x_j and x_k

The selection whether the variance-covariance matrix S or correlation matrix R to be used for the Principal Components Analysis is mainly based on the nature of the data. If the involved data are all in the same unit, same digit, and same magnitude order or the data are in percentage, and then the analysis can be carried out on variance-covariance matrix. If the data or measurements are in different unit like pressure, temperature, flow rate, energy and so on, or the data are in different magnitude order, then the Principal Components Analysis shall be performed on the correlation matrix (Marriott, 1974).

2.2.2 Determine the Eigenvalues and Eigenvector of the Matrix

The PCA is performed either on variance-covariance matrix or correlation matrix. Let A represents either variance-covariance matrix S or correlation matrix R. In order to determine *eigenvalues* and *eigenvectors* of A, the characteristic equation is form based on the Equation 2.17. Equation 2.17 has to follow the criteria as shown in Equation 2.18. However, the function equation can be expressed in the polynomial form as shown in Equation 2.19 and is called the characteristic equation of matrix A:

$$f(\lambda) = |A - \lambda I| = \begin{vmatrix} a_{1,1} - \lambda & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} - \lambda & \dots & a_{2,n} \\ \dots & \dots & \dots & \dots \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} - \lambda \end{vmatrix} \quad (2.20)$$

$$f(\lambda) = |A - \lambda I| = 0 \quad (2.21)$$

$$c_0 \lambda^n + c_1 \lambda^{n-1} + \dots + c_{n-1} \lambda + c_n = 0 \quad (2.22)$$

The order of the characteristic equation equals to the number of variables involved in the Principal Components Analysis. The solutions of the characteristic equation produces *characteristic root*, *latent root*, or *eigenvalues*, $\lambda_1, \lambda_2, \dots, \lambda_m$. The *characteristic vectors*, or *eigenvectors* can be obtained from the following equations:

$$|A - \lambda I| w_i = 0 \quad (2.23)$$

$$v_i = \frac{w_i}{\sqrt{w_i' w_i}} \quad (2.24)$$

Vectors w_1, w_2, \dots, w_n are the un-normalized *eigenvectors*. v_1, v_2, \dots, v_m are normalized *eigenvectors*. V is the *eigenvectors* matrix of variance-covariance matrix or correlation matrix. Matrix V is ortho-normal that is: $v_i^T v_i = 1$ and $v_i^T v_k = 0$. This means that the newly formed variables or Principal Components are independent with each other or in another word, orthogonal.

2.2.3 Number of Principal Components Required

Eigenvalue is a measure of the variation explained by an *eigenvector*. Each *eigenvector*, v_i associates with its *eigenvalue*, λ_i . The first Principal Component explains the largest variation of the original data. The m^{th} -Principal Component explains the smallest variation. If the variance-covariance matrix is used for Principal Components Analysis, the total *eigenvalues* equals to the total variance of the original data matrix. If the correlation matrix is used, the total *eigenvalues* equals to the number of the variables involved because the variance for each variable equals to one after standardization. The percentage of the variation to be explained by a particular *eigenvector* is determined via Equation 2.22 while the accumulated variation to be explained until j^{th} Principal Component is calculated via Equation 2.23:

Percentage of original variation explained by principal component i

$$= \frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_m} \times 100\% \quad (2.25)$$

Percentage of original variation explained by first j principal components

$$= \frac{\lambda_1 + \lambda_2 + \dots + \lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_j + \dots + \lambda_m} \quad (2.26)$$

When PCA is applied to reduce the original variables' dimensions, the percentage of the variation to be explained by the Principal Components can be fixed by the user. However, there are several suggestions on determining the number of Principal Components to be retained (Sharma, 1996):

- i) In the case of standardized data, retain only those components whose *eigenvalues* are greater than one. This is referred to as the *eigenvalues-greater-than-one* rule.
- ii) Plot the percent of the variance accounted by each principal component and look for an elbow. The plot is referred to scree plot, which is shown in Figure 2.4.
- iii) Retain components that are statistically significant

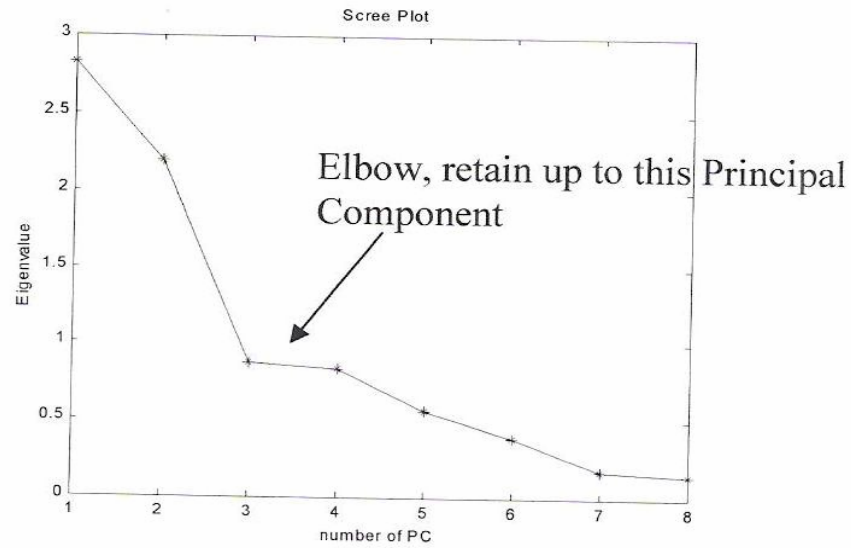


Figure 2.4 Scree Plot

If more than half numbers of total Principal Components are required to account for a reasonable variation, then the data cannot be summarized very well with a few components and shall stick to the original data itself (Cliff, 1987). The summary procedures of Principal Components Analysis are presented in Figure 2.5.

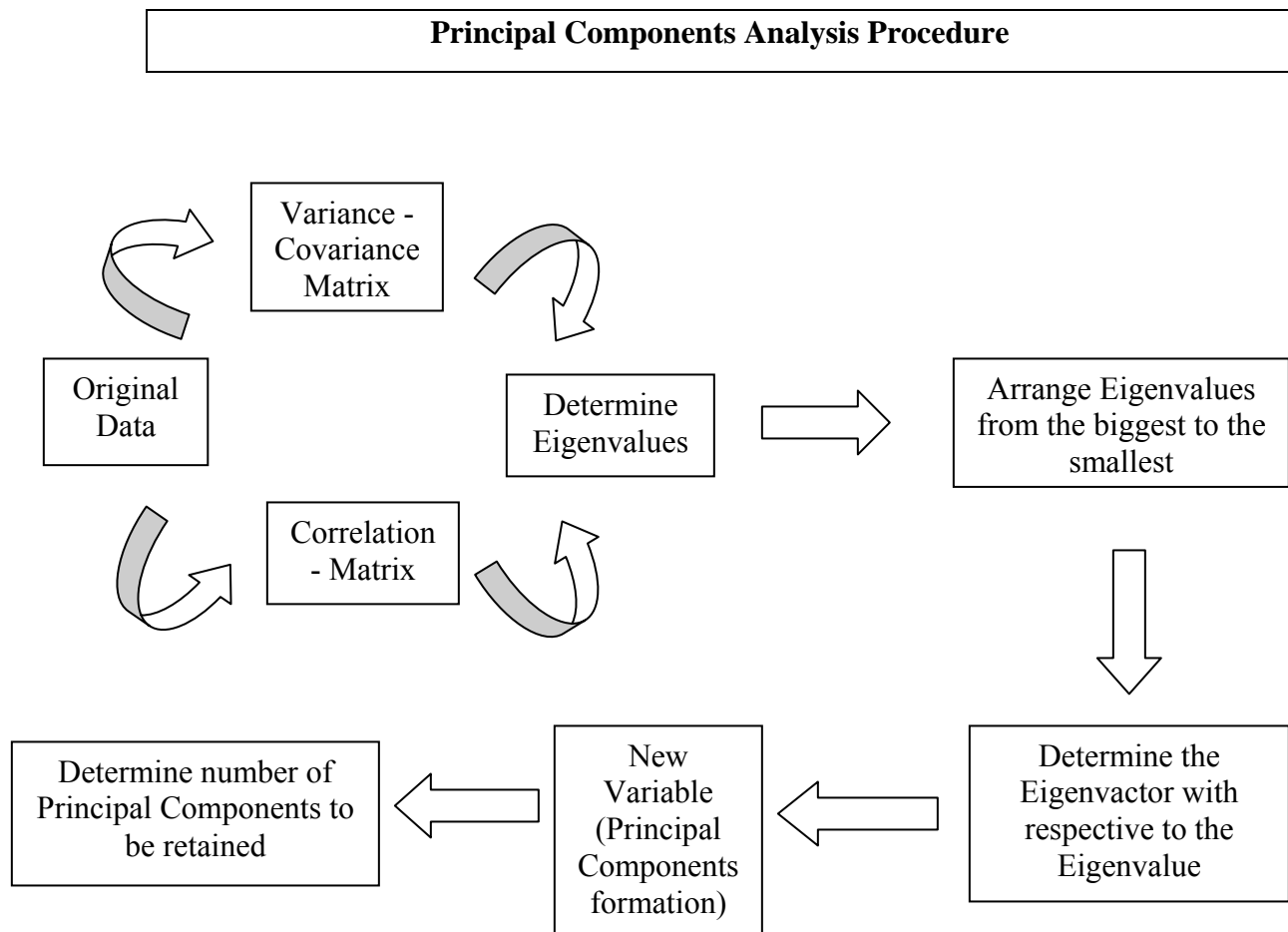


Figure 2.5 Principal Component Analysis Procedures

CHAPTER III

LITERATURE REVIEW

3.1 Gas Density Measurement

E. Sanwani presents and interprets results of experimental measurements of the spatial gas hold-up distribution in a 3 m³ glass rectangular flotation cell at the JKMRC using two different techniques. The gas hold-up device with the capturing technique was developed at the JKMRC and has been used widely in the P9 project1 while the one with conductivity technique was developed at the CSIRO Thermal and Fluids Engineering laboratory at Highett, Victoria, Australia. Measurements were conducted at more than 64 locations in the cell to determine the local gas hold-up distribution in the cell. Since the measurements using the two techniques were conducted at the same locations, the results may be compared with each other. The results indicate that the gas hold-up varies widely inside the flotation cell. The gas hold-up distributions measured by the two techniques are relatively similar except in some locations which can be reasonably explained. [3].

On the base of modern probability approached by the Derevich I.V.,the theoretical model of turbulent relative motion of particles in the turbulent flow is developed. Closed equation for probability density function of coordinates and velocities of two particles in turbulent flow is obtained. The system of equations for balance of mass, averaged velocities and intensities of turbulent chaotic motion of particles with account of correlated motion of particles are deduced. The closed expressions for intensity of relative chaotic motion between particles are obtained on

the base of probability density function of particles displacement with correlation effects. The correlation functions, intensity of relative turbulent motion and relative diffusion coefficients of particles are numerically investigated. The calculation results are compared with data of large eddy simulations. The results of calculation intensity of droplets relative motion in atmospheric conditions are presented [4].

Masahiro proposed that the airflow standard system in Japan for the middle range (5–1000 m³/h) has been renewed since 2002. The constant volume tank to realize the standard was replaced to have better performance at higher pressure, although its principle is exactly the same as its predecessor. Otherwise, most hardware for the system is unchanged, however the software for controlling the measurements is improved to give the capability for tele-calibration on a unique basis. The facility also has a unique time measurement system based on analog-to-digital (A/D) conversion to analyze instantaneous behaviors of flowmeters. This paper describes the outline of the renewed standard system together with some results from inter-laboratory comparison tests, in which very good consistency of the renewed system with other facilities, including its own predecessor, are shown [5].

Designing natural gas pipelines to safely and efficiently handle unsteady flows, requires knowledge of pressure drop, flowrate and temperature distribution throughout the system. E. Tentis were covered the accurate prediction of these parameters is essential in order to achieve optimum cumulative deliverability, and safe and reliable operation. An Adaptive Method of Lines algorithm is formulated for the solution of Euler system of equations, which fully simulates slow and fast transients. Two test cases present the improvement of the numerical solution from grid adaptation. Good results are obtained both for slow and fast transients simulations proving that the suggested numerical procedure is appropriate for such predictions [6].

A generalized fluidized bed reactor model which covers the three fluidization flow regimes most commonly encountered in industry (bubbling, turbulent and fast fluidization) is proposed. In this paper, I. A. Abba extend the model to cases where the volumetric gas flow changes appreciably due to variations in molar flow, pressure and temperature. For the air-based oxy-chlorination process as a case study, it is shown that the volume change affects both the hydrodynamics and the reactor

performance. Because the reactions are rapid, almost complete conversion of ethylene is attained immediately above the distributor resulting in 25% reduction in volumetric flowrate. Using the probabilistic averaging technique, the model tracks the probability of being in the bubbling, turbulent and fast fluidization regimes along the reactor height. The impacts of temperature and pressure variations are also examined. The variable density model gives predictions which compare well with commercial data; ignoring density variations leads to significant under prediction [7].

Oystein L. B. proposed that Multiphase meters (MPMs) are used by the petroleum industry for measurement of individual flowrates of water, gas and oil flowing through a pipeline. Different meters are available on the market, most of them detect water by a measurement of the dielectric properties of the fluid mixture. To be able to deduce water content from a measurement of the mixture permittivity, the dielectric properties of each of the constituents must be known in advance. While the oil and gas permittivity is known from information about their respective density, water permittivity depends on the salt content. If water is the continuous phase in a multiphase flow, it is crucial to know the water salinity for a good performance of MPMs. A dedicated water salinity sensor has been developed, the purpose being to detect water conductivity and salinity in a water continuous water/oil/gas/ mixture. The sensor is based on microwave spectroscopy and measures differential attenuation and phase-shift of electromagnetic radiation propagating through the mixture. This paper presents the basic principles, physical models and main design of the salinity sensor as well as experimental results obtained from flow loop test [8].

Ultrasonic gas flowmeters typically use narrowband piezoelectric transducer arrangements for interrogating the flow of gas in a pipe. I. J. O'Sullivan and W. M. D. Wright proposed, the suitability of broadband electrostatic transducers operating at frequencies of up to 1 MHz for ultrasonic measurement of gas flow has been investigated. The transit time method of ultrasonic gas flow measurement was adopted and experiments were carried out using a laboratory test rig capable of producing a range of gas flowrates up to 17.5 m/s. The test rig also allowed easy interchange of different prototype flowmetering sections. Times of flight of ultrasonic waves interrogating the gas flow were measured using separate send/receive electrostatic transducer arrangements. Two flowmeter configurations

were considered. The first interrogated the flow at 45° in contra-propagating upstream and downstream directions. The second consisted of an upstream interrogation at 45° to the gas flow and an interrogation made normal to the flow direction. k factors correlating the fluid velocity along the ultrasonic path with the mean fluid velocity in the pipe were calculated using experimental ultrasonic data and anemometer measurements. All transducer configurations were numerically modelled using the computational fluid dynamics software package FLOTRAN© (ANSYS Inc.). Theoretical gas flow velocities for both transducer arrangements were subsequently compared with experimental values and found to be in excellent agreement. A flow-dependent frequency shift of the received ultrasonic signals was also observed simultaneously with the transit time measurement [9].

R. B. Thorpe, G. M. Evans, K. Zhang and P. M. Machniewski were proposed the dispersion of bubbles into down-flowing liquids is often encountered in a number of industrial applications involving pipe flow, bubble columns and loop reactors. Usually a gas horizontal sparging device is used to generate bubbles that are carried downward with the bulk liquid flow. At low gas flowrates discrete bubbles are formed. However, at higher gas flowrates a ventilated cavity attached to the sparger is formed. For downward pipe flow the liquid forms an annular jet, which entrains gas into the recirculation region immediately beneath the ventilated cavity. The rate of gas entrainment and the size of the bubbles produced is determined by the pipe diameter, liquid and gas volumetric flowrates and the strength of the recirculation region below the base of the ventilated cavity. In this study a model was developed to predict the liquid velocity field and bubble breakup in the recirculation region. The velocity profile was modeled using the potential flow solution of the Hill's vortex, where the strength of the vortex was assumed to be directly proportional to the velocity of the annular wall jet. The proportionality constant was found to be 0.38, based on predictions obtained using the commercial code CFX. The CFX velocity profile predictions for the central part of the recirculation region were very similar to the Hill's vortex velocity profile. Bubble breakup was modeled using a critical Weber number concept, based on the predicted velocity profile within the recirculation region. It was found that the prediction of bubble size was in general agreement with experimental observations when a critical Weber number of 4.7 was assumed. A digital high-speed video was used to observe the liquid and bubble motion at the base

of the ventilated cavity. The video was used to obtain estimates of the recirculating liquid flow velocity, which compared reasonably well with predictions based on the Hill's vortex model. The video evidence also highlighted the unsteady nature of the flow, particularly the actual gas entrainment process, and possible reasons for this behavior are presented [10].

The effect of gas entrainment in oil on the performance of a range of single-phase flowmeters has been investigated experimentally using the National Standard Multiphase Flow facilities at NEL by A. F. Skea and A. R. W. Hall. The flowmeters tested were 4-inch and 2-inch positive displacement, venturi, helicoidal and flat-bladed turbine meters and 2-inch U-tube and 1.5-inch straight tube Coriolis meters. The flowmeters were tested in oil flow with gas fractions up to 15% by volume. The aim of the project was to quantify the effect of second-phase fluid components on the basic uncertainty of a range of single-phase flowmeters and, as a consequence, identify which generic types of single-phase flowmeter were most suitable in applications where such components may be present. These tests have provided evidence of the suitability of particular flowmeters for two-component flow applications. Comparisons have been made between generic type and size of flowmeter. At low gas fractions, the positive displacement and venturi flowmeters were more accurate than the other meters and estimated the total flowrate to within $\pm 2\%$. Over 9% gas fraction, there was an improvement to the response from some of the flowmeters with increasing gas fractions. This was considered to be indicative of improved mixing in the flow [11].

The paper reports from Bodo M. et. al., is research project initiated in the PTB for systematic investigations of installation effects and for finding efficient ways to minimise these effects. The project covers the design of an automated test facility using a laser Doppler anemometer, the measurement of velocity profiles downstream of several pipe configurations and flow conditioners as well as the measurement of the error shift of a turbine meter due to the disturbed velocity profiles. In this paper the test facility for investigations of installation effects will be described and the relation between pipe configuration and disturbed flow profile will be shown for a wide variety of pipe configurations and flow conditioners. More than 150 velocity distributions have been determined for different pipe configurations at several

flowrates, and data have been collected to describe the corresponding flowmeter's behaviour. Some typical velocity profiles for the most common pipe elements and flow conditioners are shown [12].

3.2 Principal Component Analysis

Dorsey, A.W. and Lee, J.H. , propose batch-to-batch prediction models based on principal component analysis. The modeling purposely to capture correlations among the variables within each individual batch as well as those of successive batches, and uses these models to predict the future behavior of the relevant variables. The main strength of this project is it allows for direct use of previous batch information together with current batch information for predicting the performance of key batch variables under consideration [13].

Principal Components Analysis (Jackson, 1980; Geladi and Kowalski, 1986; Wold *et al.*, 1987; Jackson, 1991; Johnson and Wichern 1992) can be described via using mathematical representation and graphical representation. From the mathematical point of view, Principal Components Analysis is a multivariate technique, which can transform a set of original variables x_1, x_2, \dots, x_m to a set of new variables P_1, P_2, \dots, P_m . Actually, the newly formed variables are called Principal Components (PC), which are the linear combinations of the original variables [14].

Besides, Sasongko, S.B. (2000), also proposes an integral MSPC which consist of fault detection, fault isolation and fault diagnosis. The fault detection and the isolation utilise the multivariate analysis and the control chart methods. Three types of multivariate analysis are investigated: (i) the series multi block Principal Component Analysis (SBPCA), (ii) the parallel multi block PCA (PB-PCA) and the parallel multi block PCA T^2 Hotelling (PB-PCAT). The results of this study demonstrated the potentials of multivariate statistical process control in solving fault detection and diagnosis problems for multivariable and multi equipment system [15].

Recently, Wang, S.W.(2004) developed a statistic method based on PCA (Principal Component Analysis) to detect the sensor fault occurred in AHU. And he used the contribution plots combined with some rules to isolate the fault source. To further develop this method, this paper presents an FTC method for outdoor air and AHU supply air temperature control of VAV systems to satisfy ASHRAE Standard 62-2001. PCA method, joint angle method and compensatory reconstruction are used to detect, isolate, and reconstruct the fault, respectively. Compared with conventional FDD methods, the FTC method presented in this paper has its unique advantages. First of all, it fits complicated systems especially the HVAC systems with many sensors and control variables. In addition, no need to build an accurate mathematic model, which after all is still difficult to set up in HVAC systems. Last but not least, the method has the achievement of fast, accurate detection and isolation. They are tested and evaluated under various conditions online on a centralized VAV air-conditioning system simulated using detailed HVAC dynamic models [16]

Other classical techniques from Yeung, K. Y. (2001), such as principal component analysis (PCA), have also been applied to analyze gene expression data. Using different data analysis techniques and different clustering algorithms to analyze the same data set can lead to very different conclusions. Their goal is to study the effectiveness of principal components (PC's) in capturing cluster structure. In other words, they empirically compared the quality of clusters obtained from the original data set to the quality of clusters obtained from clustering the PC's using both real and synthetic gene expression data sets. From empirical study showed that clustering with the PC's instead of the original variables does not necessarily improve, and often degrade, cluster quality. In particular, the first few PC's (which contain most of the variation in the data) do not necessarily capture most of the cluster structure. It is also showed that clustering with PC's has different impact on different algorithms and different similarity metrics. Overall, they would not recommend PCA before clustering except in special circumstances.

In addition, Ramalingam, S. (2005) proposed data envelopment analysis (DEA), a popular linear programming technique is useful to rate comparatively operational efficiency of decision making units (DMU) based on their deterministic (not necessarily stochastic) input–output data. Only when the input–output data are

stochastic (preferably distributed as a multivariate Gaussian), a statistical technique called principal component analysis (PCA) could alternatively be used for the same purpose of rating DMU. Because of these choices, research interest has been growing among statisticians and mathematical programmers to explore benefits versus disadvantages of using one technique over the other. Yet, the duality between DEA and PCA has not been fully understood. This article is devoted to investigate their complementarities. With an expectation that an integration of both techniques would offer the best of DEA and PCA, several integration methods have been suggested in the literature. In these methods, ratio of two Gaussian random variables is involved and this creates a flaw. The ratio is Cauchy distributed and not Gaussian distributed. Neither mean nor dispersion exists in Cauchy distribution. To overcome this flaw of trapping into a Cauchy distribution, a novel method of integrating DEA and PCA, as it is proposed and demonstrated in this article, would enrich the validity of the integration. A medical example is considered for illustration. In the medical example, 45 countries are rated with respect to their survival rate from melanoma cancer among men and among women as output data variable and data on location latitude, ozone thickness, ultraviolet rays of type A and type B as input data variables. Firstly, DEA, secondly PCA are separately applied and then thirdly integrated approach of this article is tried on data. The results are compared and commented with a few concluding thoughts [17].

Principal Component Principles (PCA) based algorithm to extract cracks in concrete bridge decks for the purpose of automating inspection is presented by Ikhlas A.Q et al.,(2006). PCA will be used to identify clusters using a database of bridge images. Results from three different PCA approaches are presented in this work. The first approach employs PCA by itself on raw data. In the second approach, a linear structure modeling is implemented prior to PCA processing in an effort to enhance the results since cracks can be detected as linear structures. Several convolution-processes with masks designed to identify linear structure in the data are used. In both cases, attempts to detect cracks in a global framework were used. The third approach, on the other hand, used local information (neighborhoods) instead of global. That is, each image is segmented into small blocks where each block is processed as an individual entity. Experimental results show enhancement in the local detection with linear modeling over the global [18].

Carlos G. et al., (2006) proposed to analyse the magnetoencephalogram (MEG) background activity in patients with Alzheimer's disease (AD) using the Lempel–Ziv (LZ) complexity. This non-linear method measures the complexity of finite sequences and is related to the number of distinct substrings and the rate of their occurrence along the sequence. The MEGs were recorded with a 148-channel whole-head magnetometer (MAGNES 2500 WH, 4D Neuroimaging) in 21 patients with AD and in 21 age-matched control subjects. Artefact-free epochs were selected for complexity analysis. Results showed that MEG signals from AD patients had lower complexity than control subjects' MEGs and the differences were statistically significant ($p < 0.01$). In order to reduce the dimension of the LZ complexity results, a Principal Components Analysis (PCA) was applied, and only the first principal component was retained. The first component score from PCA was graphically analysed using a box plot and a receiver-operating characteristic (ROC) curve. A specificity of 85.71%, a sensitivity of 80.95% and an area under the ROC curve of 0.9002 were obtained. These preliminary results suggest that cognitive dysfunction in AD is associated with a decreased LZ complexity in the MEG signals [19].

Producing the grain with equilibrium moisture content is an important process control objective. Xueqiang *et. al*, (2005) achieving this objective and it can be very difficult in grain drying process because of its multi-variables, nonlinearity and long delay. In this paper, a control approach based on principal component analysis (PCA) is presented to achieve this objective. A PCA model which incorporates time lagged variables is used, and the control objective is expressed in the score space of this PCA model. A controller is designed in the model predictive control framework, and it is used to control the equivalent score space representation of the process. The score predictive model for the model predictive control algorithm is built using neural network partial least squares (NNPLS). The process control system with NNPLS was tested on a commercial mixed-flow dryer and showed excellent accuracy and stability [20].

A straight line detection algorithm is presented by Yun-Seok L. *et.al*, (2006). The algorithm separates row and column edges from edge image using their primitive shapes. The edges are labeled, and the principal component analysis (PCA)

is performed for each labeled edges. With the principal components, the algorithm detects straight lines and their orientations, which is useful for various intensive applications. Our algorithm overcomes the disadvantages of Hough transform (HT) and other algorithms, i.e. unknown grouping of collinear lines, complexity and local ambiguities. The experimental results show the efficiency of our algorithm [21].

On the other hand, in this study Kemal P. et al, (2006), have detected on lung cancer using principles component analysis (PCA), fuzzy weighting pre-processing and artificial immune recognition system (AIRS). lung cancers are cancers that begin in the lungs. Other types of cancers may spread to the lungs from other organs. However, these are not lung cancers because they did not start in the lungs. It is evident that usage of machine learning methods in disease diagnosis has been increasing gradually. In this study, diagnosis of lung cancer, which is a very common and important disease, was conducted with such a machine learning system. The approach system has three stages. First, dimension of lung cancer dataset that has 57 features is reduced to four features using principles component analysis. Second, a new weighting scheme based on fuzzy weighting pre-processing was utilized as a pre-processing step before the main classifier. Third, artificial immune recognition system was our used classifier. They took the lung cancer dataset used in our study from the UCI machine learning database. The obtained classification accuracy of our system was 100% and it was very promising with regard to the other classification applications in literature for this problem [22].

Weixiang S. et. al, (2006) proposed analyzing the flaws of conventional fault diagnosis methods, data mining technology is introduced to fault diagnosis field, and a new method based on C4.5 decision tree and Principal Component Analysis (PCA) is proposed. In this method, PCA is used to reduce features after data collection, preprocessing and feature extraction. Then, C4.5 is trained by using the samples to generate a decision tree model with diagnosis knowledge. At last the tree model is used to make diagnosis analysis. To validate the method proposed, six kinds of running states (normal or without any defect, unbalance, rotor radial rub, oil whirl, shaft crack and a simultaneous state of unbalance and radial rub), are simulated on Bently Rotor Kit RK4 to test C4.5 and PCA-based method and back-propagation

neural network (BPNN). The result shows that C4.5 and PCA-based diagnosis method has higher accuracy and needs less training time than BPNN [23].

CHAPTER IV

METHODOLOGY

4.0 Introduction

To ensure the research could be done in time as intended, the methodology is divided into two parts which are case study and method algorithm. During the case study, the AFPT 921 Plant (Air Flow Pressure Temperature) is used to gather the inputs data. Then the inputs data are manipulated to make the output one. In method algorithm, there will be operational framework of the research.

4.1 Case Study of AFPT 921

AFPT 921 Plant is a process control training system which is related to the study of air (gas), flow rate, pressure and temperature in the pipeline and the tank as well. The process measurement and control system as follows:

- a) Temperature Measurement & Control
 - i) PID Temperature Control, Single loop
 - ii) ON/OFF Temperature Control
 - iii) PID Temperature Auto-Selector Control

b) Flowrate Measurement & Control

- i) Air Flow Measurement in a Pipeline-Volumetric and Mass Flowrate
- ii) Uncompensated gas flow measurement, $F_v = K_1 \cdot \sqrt{h}$
- iii) Compensated gas flow measurement

$$\text{Mass Flowrate, } F_m = k_m \cdot \frac{\sqrt{(hp)}}{\sqrt{T}}$$

$$\text{Volumetric Flowrate, STP or NTP based, } F_{vb} = k_{vb} \cdot \frac{\sqrt{(hp)}}{\sqrt{T}}$$

$$\text{Volumetric Flowrate, at flowing condition, } F_v = k_v \cdot \frac{\sqrt{(hp)}}{\sqrt{P}}$$

- v) Absolute engineering unit of Pressure and Temperature
- vi) Variable area flow meter for gas
- vii) PID Flow Control
- viii) Low Flow Detection and Control

c) Pressure Measurement & Control

- i) ON/OFF Air Pressure Control
- ii) PID Air Pressure Control, Single Loop, Single Capacity or Pipeline
- iii) PID Air Pressure Control, Single Loop, Multicapacities
- iv) PID Air Pressure Control, Cascade

4.2 Methodology Framework

The Figure 3.1 below shows the steps to develop this project. Among that, the development of gas density measurement using Matlab is the crucial part. It involves understanding the concept of multivariate analysis and PCA technique plus writing source data.

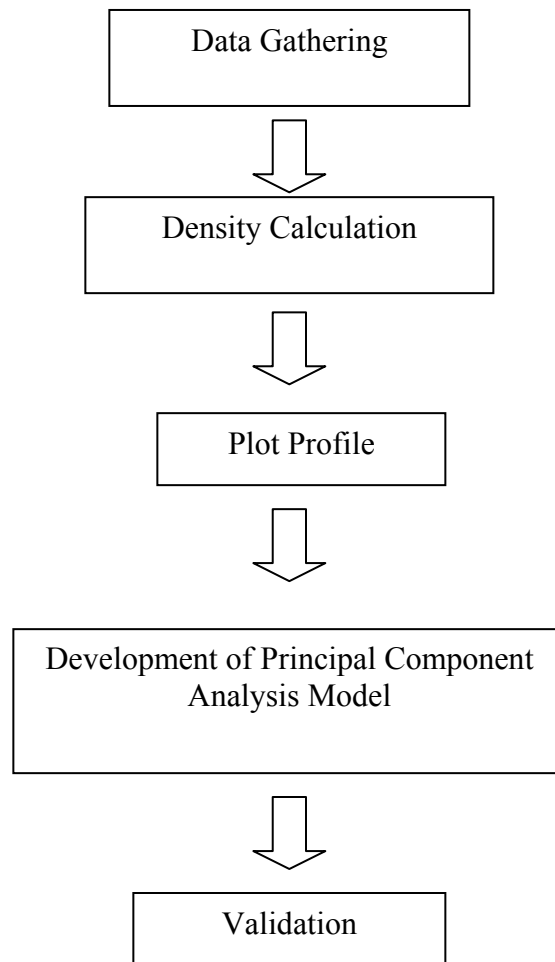


Figure 4.1 Flowcharts for Methodology

4.2.1 Data Gathering

Data is gathered through the experiment of AFPT Plant in which two input variables were measured. The data is gathered in such a way is shows significant stability as the variables change with time from one setting value to another.

4.2.2 Density Calculation

As the density is to be determined, the input variables were transferred to the excel so that the density can be calculated. The calculation method is done by manipulating the ideal gas law while the input variables are used simultaneously to the equation.

4.2.3 Plot Profile

Variables profile is the graph illustration of correlation between density and both of input variables. The graph is purposely used to show the dynamic behavior between input variables and output variable.

4.2.3 Model Development

The PCA technique is used with MATLAB to develop the measurement model:

- i) Get some data
- ii) Subtract the mean
- iii) Calculate the covariance matrix

- iv) Calculate the eigenvectors and eigenvalues of the covariance
- v) Choosing components and forming a feature vector
- vi) Deriving the new data set

4.2.4 Validation

The validation process is the process used to see the validity of the final equation respective to the theory one. The result from the real process will be compared to the result in MATLAB to validate the model. The percentage error is calculated by comparing the manual calculation (excel calculation) with the value of PCA technique.

CHAPTER V

RESULTS AND DISCUSSION

5.0 Data Gathering

Figure 5.1 and Figure 5.2 show the data that gathered from the experiment of AFPT 921 Plant.

Table 5.1 Temperatures in Pipeline (Experiment Data)

TIME (mint)	SET POINT (°C)	PROCESS TEMPERATURE (°C)	PROCESS TEMPERATURE (K)
Steady State	60	52.93584	326.08584
2	60	54.99104	328.14104
4	60	56.98901	330.13901
6	60	58.90116	332.05116
8	60	60.90143	334.05143
10	70	62.98641	336.13641
12	70	64.93421	338.08421
14	70	66.94769	340.09769
16	70	68.99101	342.14101
18	70	70.92111	344.07111
20	80	72.96864	346.11864
22	80	74.99011	348.14011
24	80	76.97012	350.12012
26	80	98.9321	372.0821
28	80	80.94112	354.09112
30	90	82.9214	356.0714
32	90	84.9711	358.1211
34	90	86.95121	360.10121
36	90	88.98143	362.13143

38	90	90.90091	364.05091
40	100	92.97101	366.12101
42	100	94.9344	368.0844
44	100	96.98323	370.13323
46	100	98.97447	372.12447
48	100	100.91134	374.06134
50	110	102.93411	376.08411
52	110	104.90013	378.05013
54	110	106.99932	380.14932
56	110	108.97421	382.12421
58	110	110.965	384.115
60	120	112.9547	386.1047
62	120	114.92101	388.07101
64	120	116.93107	390.08107
66	120	118.94441	392.09441
68	120	120.97521	394.12521
70	130	122.97012	396.12012
72	130	124.98421	398.13421
74	130	126.9321	400.0821
76	130	128.92101	402.07101
78	130	130.97667	404.12667
80	140	132.91431	406.06431
82	140	134.92121	408.07121
84	140	136.956	410.106
86	140	138.953	412.103
88	140	140.9939	414.1439
90	150	142.9985	416.1485
92	150	144.9912	418.1412
94	150	146.9701	420.1201
96	150	148.9443	422.0943
98	150	150.0172	423.1672

Table 5.2 Pressures in Pipeline (Experiment Data)

TIME (mint)	SET POINT (kg/Hr.)	FLOW RATE IN PIPELINE (kg/Hr.)	PRESSURE IN TANK (Pa)
Steady state	19	19.91088	137280.018
2	19	18.64366	128542.8861
4	19	18.65439	128616.8665
6	19	19.08296	131571.7383
8	19	18.55604	127938.7704
10	18	19.03587	131247.0657
12	18	17.90515	123451.0637
14	18	18.00588	124145.5692
16	18	18.05774	124503.1296
18	18	18.08874	124716.866
20	17	17.67984	121897.6134
22	17	16.61529	114557.8352
24	17	16.8698	116312.6114
26	17	17.02299	117368.8141
28	17	17.05637	117598.96
30	16	16.70052	115145.4725
32	16	15.90121	109634.4509
34	16	16.43775	113333.746
36	16	16.05672	110706.6494
38	16	15.94333	109924.8566
40	15	16.94721	116846.332
42	15	15.23561	105045.3227
44	15	15.30749	105540.9155
46	15	16.4309	113286.5171
48	15	16.73473	115381.341
50	14	13.68091	94326.09564
52	14	13.98668	96434.29533
54	14	13.92588	96015.09612
56	14	13.9986	96516.48044
58	14	13.91098	95912.36474
60	13	13.94734	96163.0569
62	13	12.85835	88654.77164
64	13	12.98352	89517.78421
66	13	12.99604	89604.10616
68	13	13.0318	89850.66148
70	12	13.02942	89834.25204
72	12	11.87724	81890.28918
74	12	12.03996	83012.19864
76	12	12.03222	82958.83347
78	12	12.00897	82798.53115
80	11	12.01076	82810.8727
82	11	10.80553	74501.14475
84	11	10.93845	75417.59143
86	11	10.98077	75709.37615
88	11	10.99329	75795.69809

90	10	11.01952	75976.54669
92	10	10.44492	72014.8384
94	10	10.34419	71320.33288
96	10	10.33226	71238.07882
98	10	10.31617	71127.14271

The temperature and pressure were manipulated by time to evaluate the dynamic performance. The temperature and pressure value are obtained from the pipeline setting only. The data is redundant when the process has reached the steady state.

5.1 Gas Density Measurement

The density is measured by manipulating the ideal gas law which the gas condition is assumed as ideal gas. Then the constant value is developed and simultaneously the pressure and the temperature were used to calculate the density. The constant value is developed as shown below:

We know that,

$$\rho = \frac{m}{V}$$

Perfect Gas Law : $PV = nRT$

$$n = \frac{m}{mw} \quad , \quad \frac{mw}{R} = \text{constant, } K$$

Rearrange the Perfect Gas Law

$$\frac{m}{V} = \frac{mw}{R} \frac{P}{T}$$

$R=8.314 \text{ J/mol.K}$ and $mw = 28.98 \text{ kg/mol}$

$$\rho = K \frac{P}{T} \quad , \quad K = 3.4845 \quad (5.1)$$

$$\rho = 3.4845 \frac{p}{T}$$

Table 5.3 Results of Density Calculation

TIME (mint)	PROCESS TEMPERATURE (K)	PRESSURE IN TANK (Pa)	DENSITY IN TANK (ρ)
Steady State	326.08584	137280.018	1466.951839
2	328.14104	128542.8861	1364.985272
4	330.13901	128616.8665	1357.505347
6	332.05116	131571.7383	1380.696041
8	334.05143	127938.7704	1334.532965
10	336.13641	131247.0657	1360.550023
12	338.08421	123451.0637	1272.361201
14	340.09769	124145.5692	1271.944058
16	342.14101	124503.1296	1267.989345
18	344.07111	124716.866	1263.041002
20	346.11864	121897.6134	1227.186822
22	348.14011	114557.8352	1146.598066
24	350.12012	116312.6114	1157.577846
26	372.0821	117368.8141	1099.14353
28	354.09112	117598.96	1157.254596
30	356.0714	115145.4725	1126.808834
32	358.1211	109634.4509	1066.737604
34	360.10121	113333.746	1096.6679
36	362.13143	110706.6494	1065.241202
38	364.05091	109924.8566	1052.141754
40	366.12101	116846.332	1112.066865
42	368.0844	105045.3227	994.4198313
44	370.13323	105540.9155	993.5809332
46	372.12447	113286.5171	1060.792559
48	374.06134	115381.341	1074.81378
50	376.08411	94326.09564	873.9515218
52	378.05013	96434.29533	888.8379488
54	380.14932	96015.09612	880.0873363
56	382.12421	96516.48044	880.1108836
58	384.115	95912.36474	870.0692109
60	386.1047	96163.0569	867.8479484
62	388.07101	88654.77164	796.0335707
64	390.08107	89517.78421	799.6407492
66	392.09441	89604.10616	796.3018598
68	394.12521	89850.66148	794.3785934
70	396.12012	89834.25204	790.2336575
72	398.13421	81890.28918	716.7098568
74	400.0821	83012.19864	722.9916214
76	402.07101	82958.83347	718.9527423
78	404.12667	82798.53115	713.913491
80	406.06431	82810.8727	710.6127744
82	408.07121	74501.14475	636.1616123
84	410.106	75417.59143	640.7918863
86	412.103	75709.37615	640.1538479
88	414.1439	75795.69809	637.7254621

90	416.1485	75976.54669	636.1678029
92	418.1412	72014.8384	600.1219311
94	420.1201	71320.33288	591.5348966
96	422.0943	71238.07882	588.0891679
98	423.1672	71127.14271	585.684639

5.2 Plot Profile

Figure 5.1 and Figure 5.2 show the dynamic performance profile of the density versus temperature and pressure accordingly.

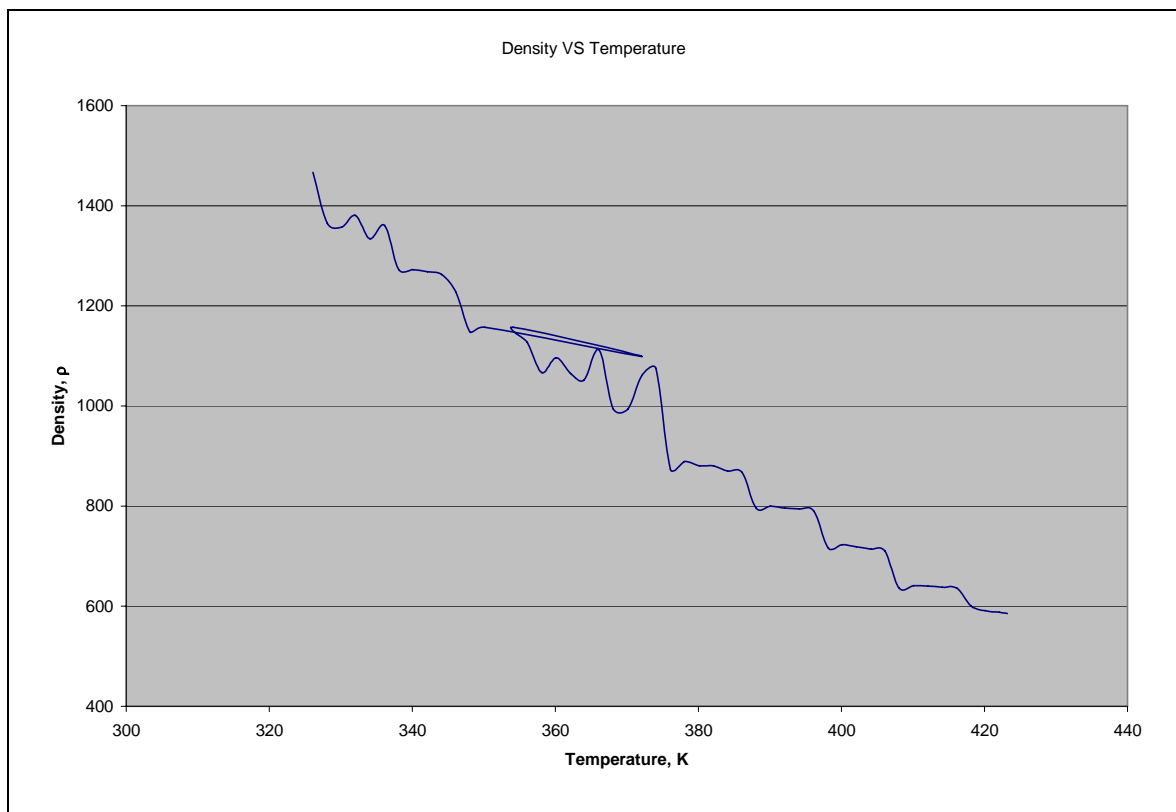


Figure 5.1 Profile Density versus Temperature

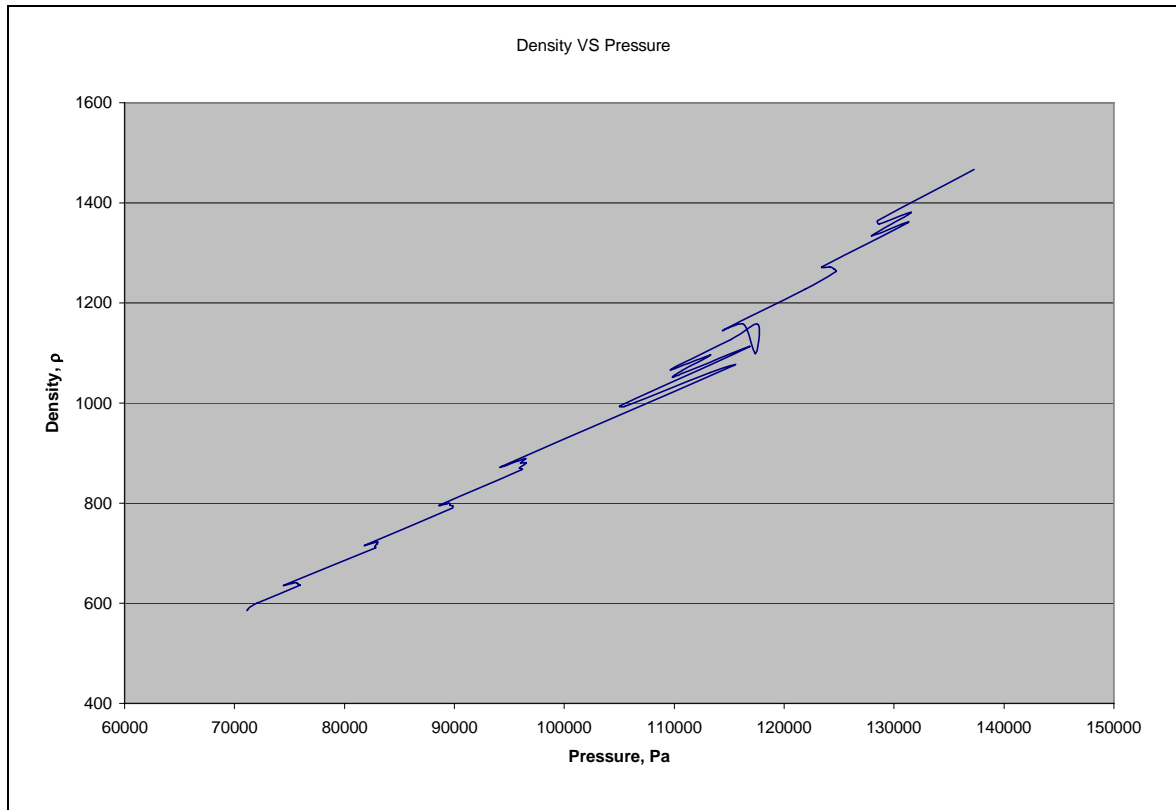


Figure 5.2 Profile Density versus Pressure

As illustrated in the first profile the density is inverse proportional as the temperature increases. It is because the increase of temperature affected the kinetic energy of the air. On the other hand, the second profile shows that the density is proportional to the pressure. It is because as the pressure increases the air molecules are getting closer to each other because of the increasing forces.

5.3 Development of Principal Component Analysis Model

The modeling was started by categorization in the M-file in which the input variables were named as follows:

x = temperature

y = pressure

```

1 x=[326.08584 328.14104 330.13901 332.05116 334.05143 336.13641 338.08421 340.09769 342.14101 344.07111 346.11864 348.1401
2
3 y=[137280.018 128542.8861 128616.8665 131571.7383 127938.7704 131247.0657 123451.0637 124145.5692 124503.1296 124716.866
4
5
6 M = mean(x)
7 M = mean(y)
8
9 x=[326.08584 328.14104 330.13901 332.05116 334.05143 336.13641 338.08421 340.09769 342.14101 344.07111 346.11864 348.1401
10
11 y=[137280.018 128542.8861 128616.8665 131571.7383 127938.7704 131247.0657 123451.0637 124145.5692 124503.1296 124716.866
12
13
14 z=cov(x,y)
15 eig(z)
16 [x,y]=eig(z)
17
18 PC = princomp(x)
19 [PC,SCORE,latent,tsquare] = princomp(x)
20
21 PC = princomp(y)
22 [PC,SCORE,latent,tsquare] = princomp(y)
23
24
25 x=[326.08584 328.14104 330.13901 332.05116 334.05143 336.13641 338.08421 340.09769 342.14101 344.07111 346.11864 348.1401
26
27 y=[137280.018 128542.8861 128616.8665 131571.7383 127938.7704 131247.0657 123451.0637 124145.5692 124503.1296 124716.866
28
29 pc = pcacov(x)
30 [pc,latent,explained] = pcacov(x)
31
32 pc = pcacov(y)
33 [pc,latent,explained] = pcacov(y)
34
35

```

Figure 5.3 The Programming Data Set in M-file

```

35
36 x=[326.08584 328.14104 330.13901 332.05116 334.05143 336.13641 338.08421 340.09769 342.14101 344.07111 346.11864
37
38 y=[137280.018 128542.8861 128616.8665 131571.7383 127938.7704 131247.0657 123451.0637 124145.5692 124503.1296 124
39
40 C = cov(x,y)
41
42 x=[326.08584 328.14104 330.13901 332.05116 334.05143 336.13641 338.08421 340.09769 342.14101 344.07111 346.11864
43
44 y=[137280.018 128542.8861 128616.8665 131571.7383 127938.7704 131247.0657 123451.0637 124145.5692 124503.1296 124
45
46 C = corrcoef(x,y)
47
48 k1=1.0000
49 k2=0.8171
50
51 P=(k1*x)+(k2*y)
52
53

```

Figure 5.4 The Programming Data Set in M-file (contd)

Figure 5.3 and Figure 5.4 show the programming data set in the M-file. The data set is coded whereby the programming codes were installed starting from naming the input variables until to correlation coefficient. The 50 of data was exported to the M-file and the reduction of the data dimension occurred once the data is being placed in the command window.

5.4 The PCA Codes

The PCA codes is the coding that being used in the modeling session which is the coding term begun with the introduction of syntax. Here is the coding for the development gas density measurement model:

i) $M = \text{mean}(A)$

$M = \text{mean}(A)$ returns the mean values of the elements along different dimensions of array. If A is a vector, $\text{mean}(A)$ returns the mean value of A . If A is a matrix, $\text{mean}(A)$ treats the columns of A as vectors, returning a row vector of mean values. If A is multidimensional array, $\text{mean}(A)$ treats the value along the first non-singleton dimension as vectors, returning an array of mean values. $M = \text{mean}(A, \text{dim})$ returns the mean value for elements along the dimension of A specified by scalar dim .

ii) $C = \text{cov}(x)$

$C = \text{cov}(x,y)$

$C = \text{cov}(x)$ where x is a vector returns the variance of the vector elements. For matrices where each row is an observation and each column a variable, $\text{cov}(x)$ is the covariance matrix. $\text{Diag}(\text{cov}(x))$ is a vector of variances for each column, and $\text{sqrt}(\text{diag}(\text{cov}))$ is a vector of standard deviations. $C = \text{cov}(x,y)$, where x and y are column vector of equal length, is equivalent to $\text{cov}([x \ y])$

- iii) $z = \text{cov}(x, y)$
 $\text{eig}(z)$
 $[x, y] = \text{eig}(z)$

Eigenvector and eigenvalues for the involved matrix

- iv) $PC = \text{princomp}(X)$
 $[PC, \text{SCORE}, \text{latent}, \text{tsquare}] = \text{princomp}(X)$

$[PC, \text{SCORE}, \text{latent}, \text{tsquare}] = \text{princomp}(X)$ takes a data matrix and returns the principal components in PC, the so-called Z-scores in SCORES, the eigenvalues of the covariance matrix of X in latent, and Hotelling's T2 statistic for each data point in tquare. The Z-scores are the data formed by transforming the original data into the space of the principal components. The values of the vector, latent, are the variance of the columns of SCORE. Hotelling's T2 is a measure of the multivariate distance of each observation from the center of data set.

- v) $S = \text{corrcoef}(x)$
 $S = \text{corrcoef}(x, y)$

$S = \text{corrcoef}(X)$ returns a matrix of correlation coefficient calculated from an input matrix whose rows are observations and whose column are variables. The matrix $S = \text{corrcoef}(X)$ is related to the covariance matrix $C = \text{cov}(X)$.

5.5 Validation

From the programming, the result that obtained from Matlab shows the value of the gas density measurement. The variables that are used in this research by using Principal Component Analysis are temperature, pressure, and density. All the results are based on the experiment of AFPT 921 Plant. Table 5.4 shows the results for gas density based on Matlab M-file programming.

Table 5.4 Results of Gas Density from Programming

PROCESS TEMPERATURE (K)	PRESSURE IN TANK (Pa)	DENSITY (programming)
326.08584	137280.018	1447.800867
328.14104	128542.8861	1378.464962
330.13901	128616.8665	1381.067426
332.05116	131571.7383	1407.123834
334.05143	127938.7704	1379.439123
336.13641	131247.0657	1408.556184
338.08421	123451.0637	1346.802851
340.09769	124145.5692	1354.491136
342.14101	124503.1296	1359.456082
344.07111	124716.866	1363.132622
346.11864	121897.6134	1342.144039
348.14011	114557.8352	1284.192181
350.12012	116312.6114	1300.510468
372.0821	117368.8141	1331.10268
354.09112	117598.96	1314.992222
356.0714	115145.4725	1296.925056
358.1211	109634.4509	1253.944198
360.10121	113333.746	1286.151249
362.13143	110706.6494	1266.715462
364.05091	109924.8566	1262.246913
366.12101	116846.332	1320.872389
368.0844	105045.3227	1226.409732
370.13323	105540.9155	1232.508051
372.12447	113286.5171	1297.788601
374.06134	115381.341	1316.842277
376.08411	94326.09564	1146.822637
378.05013	96434.29533	1166.014757
380.14932	96015.09612	1164.68867
382.12421	96516.48044	1170.760372
384.115	95912.36474	1167.814932
386.1047	96163.0569	1171.853038

388.07101	88654.77164	1112.469149
390.08107	89517.78421	1121.530885
392.09441	89604.10616	1124.249561
394.12521	89850.66148	1128.294965
396.12012	89834.25204	1130.155793
398.13421	81890.28918	1067.259763
400.0821	83012.19864	1078.374775
402.07101	82958.83347	1079.927638
404.12667	82798.53115	1080.673468
406.06431	82810.8727	1082.711951
408.07121	74501.14475	1016.820064
410.106	75417.59143	1026.34314
412.103	75709.37615	1030.724313
414.1439	75795.69809	1033.470549
416.1485	75976.54669	1036.952863
418.1412	72014.8384	1006.574445
420.1201	71320.33288	1002.87854
422.0943	71238.07882	1004.180642
423.1672	71127.14271	1004.347083

Table 5.4 represents the density measurement from the PCA technique in which the developed model is shown as follows:

$$\rho = k_1(T) + k_2(P) \quad (5.2)$$

whereby from the programming, k_1 and k_2 are the correlation coefficient between two input variables

$$k_1 = 1.0000$$

$$k_2 = 0.8171$$

Therefore the Gas Density Measurement model is

$$\rho = 1.0000 (T) + 0.8171(P) \quad (5.3)$$

Table 5.5 Comparison between Manual and Simulation

DENSITY (Ideal Gas Law)	DENSITY (programming)	Percentage of error (%)
1466.951839	1447.800867	1.305494256
1364.985272	1378.464962	-0.987533732
1357.505347	1381.067426	-1.735689591
1380.696041	1407.123834	-1.914092039
1334.532965	1379.439123	-3.364934339
1360.550023	1408.556184	-3.528437778
1272.361201	1346.802851	-5.850669601
1271.944058	1354.491136	-6.489835577
1267.989345	1359.456082	-7.213525678
1263.041002	1363.132622	-7.924653265
1227.186822	1342.144039	-9.367540047
1146.598066	1284.192181	-12.0002047
1157.577846	1300.510468	-12.34756025
1099.14353	1331.10268	-21.10362693
1157.254596	1314.992222	-13.63033049
1126.808834	1296.925056	-15.09716794
1066.737604	1253.944198	-17.54945108
1096.6679	1286.151249	-17.2780975
1065.241202	1266.715462	-18.91348735
1052.141754	1262.246913	-19.96928249
1112.066865	1320.872389	-18.77634615
994.4198313	1226.409732	-23.32917078
993.5809332	1232.508051	-24.04707154
1060.792559	1297.788601	-22.34141256
1074.81378	1316.842277	-22.51817957
873.9515218	1146.822637	-31.22268323
888.8379488	1166.014757	-31.18417801
880.0873363	1164.68867	-32.33785125
880.1108836	1170.760372	-33.02418977
870.0692109	1167.814932	-34.22092373
867.8479484	1171.853038	-35.02976416
796.0335707	1112.469149	-39.75153686
799.6407492	1121.530885	-40.25434373
796.3018598	1124.249561	-41.18384218
794.3785934	1128.294965	-42.03491564
790.2336575	1130.155793	-43.0153958
716.7098568	1067.259763	-48.91099276
722.9916214	1078.374775	-49.15453279
718.9527423	1079.927638	-50.20843158
713.913491	1080.673468	-51.37316799
710.6127744	1082.711951	-52.36314207
636.1616123	1016.820064	-59.83675285
640.7918863	1026.34314	-60.16793626
640.1538479	1030.724313	-61.01196867
637.7254621	1033.470549	-62.05571369
636.1678029	1036.952863	-62.99989693

600.1219311	1006.574445	-67.728322
591.5348966	1002.87854	-69.53835619
588.0891679	1004.180642	-70.7531267
585.684639	1004.347083	-71.48257204

The value of gas density is quite higher because the PCA technique that used in Matlab purposely to reduce the dimension of many variables. Since this research only involved two variables, the correlation coefficient obtained from PCA cannot implements all steps that stipulate in the PCA technique itself.

Both of the results for the gas density from experiment (manual) data and the simulation data were compared to see the differences between both of them. The error between the manual data and the experiment data are calculated based on the following equation:

$$\text{Error} = \left| \frac{\text{manual-programming}}{\text{manual}} \right| \times 100\% \quad (5.3)$$

5.6 Analysis and Discussion

M-file in Matlab provides and advantages to illustrate the principal component analysis. Since it is not easy task and calculating manually will become tedious when there is more variables involved, M-file helps a lot to illustrate the principal component.

After gaining data from the Microsoft Excel and from the developed model, both results were compared to validate the model. From Table 5.5, it shows that an error occurred when both results were compared. From this comparison, it was found that the best configuration is when the both of inputs are at early stage of the process.

CHAPTER VI

CONCLUSION AND RECOMMENDATION

6.0 Conclusions

At the end of this project, it was covered the objectives that have been stated at the early stage of this research which are to model the relationship between primary output (density) and inputs (temperature and pressure) using Principal Component Analysis and also to develop a program to measure the density from AFPT 921 Plant.

6.2 Recommendation

This research hopefully relies to the gas density measurement on any real plant as a tremendous solution to correlates the involved process variables. Optimistically, this research also hopefully can be expanded by the future students in order to emphasize our knowledgeable sources.

REFERENCES

- [1] Alexey P., Oxana R. and , Agnar H. (2006). *Process control and optimization with simple interval calculation method*. Chemometrics and Intelligent Laboratory Systems, Volume 81, Issue 2, 15 April 2006,165-179.

- [2] Ali R. M. and Abdulkadir G. (2006). *Materials & Design*. Statistical process control in machining, a case study for machine tool capability and process capability , Volume 27, Issue 5, , Pages 364-372.

- [3] Deming, W. E. (1982), *Out of the Crisis*, Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.

- [4] Deming, W. E., *Quality*, 1(1982). *Productivity and Competitive Position*, Massachusetts Institute of Technology. Center for Advanced Engineering Study, Cambridge.

- [5] Deming, W. E. (1982). *Quality, Productivity, and Competitive Position*, Massachusetts Institute of Technology,Center for Advanced Engineering Study, Cambridge.

- [6] Deborah F. C, Christopher W. Z., and Mary L.W. (2005). *Environmental statistical process control using an augmented neural network classification approach*. European Journal of Operational Research.

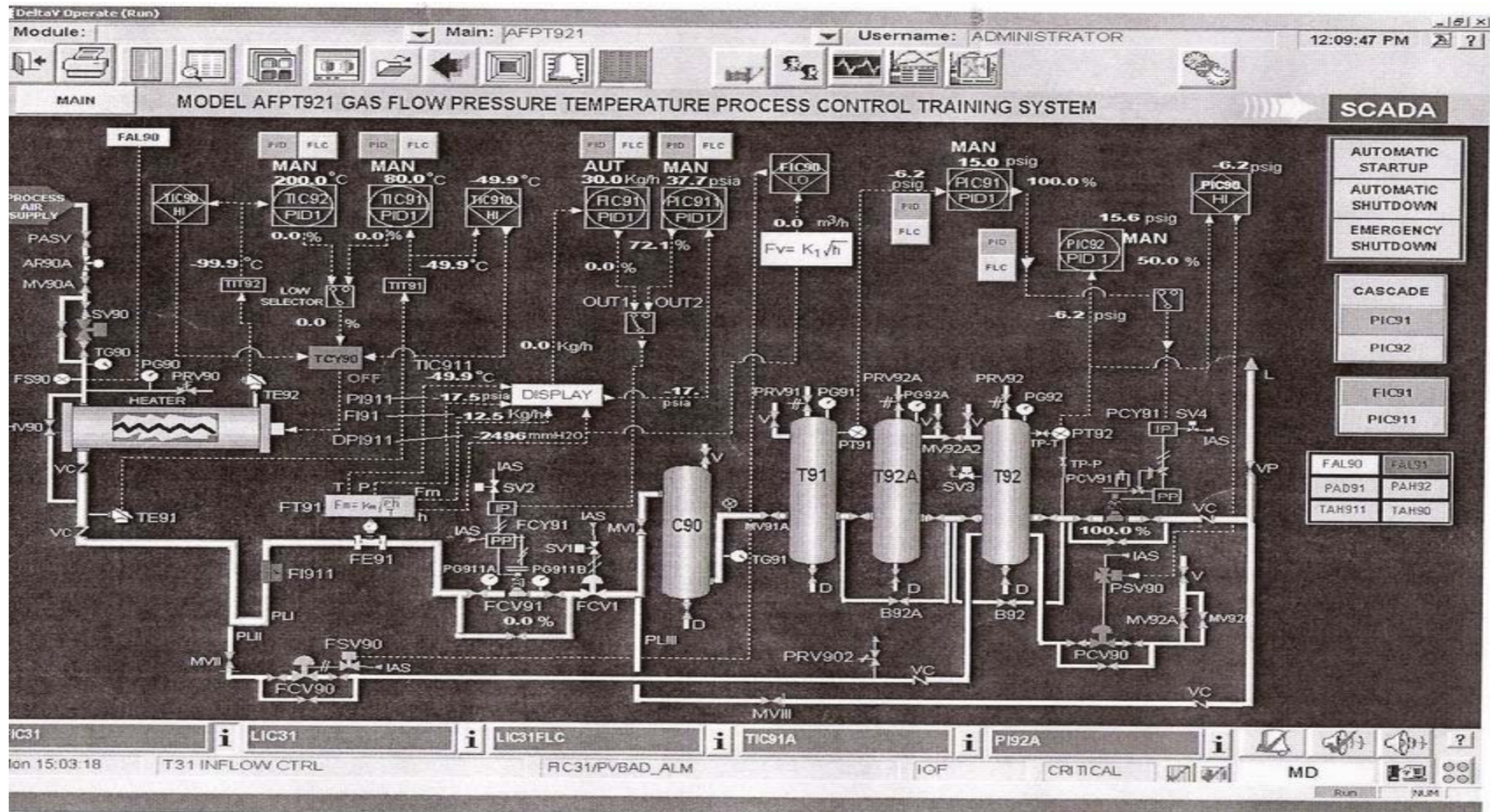
- [7] Francisco A., Carlos J. G. (2005) Design and optimization of EWMA control charts for in-control, indifference, and out-of-control regions.
- [8] Henry R. Neave and Donald J. Wheeler (1996), *Shewhart's Charts and the Probability Approach*.
- [9] Kourti, T., Lee, J. and Macgregor, J. F. (1996), *Computers chemical engineering* Experiences with industrial applications of projection methods for multivariate statistical process control.20;745-750.
- [11] L.N. VanBrackle, M.R. Reynolds, (1997). EWMA and CUSUM control charts in the presence of correlation, *Communications in Statistics—Simulation and Computation* 26, 979–1008.
- [12] Montgomery, D.C., (1992), *IIE Transactions* The use of statistical process control and design of experiments in product and process improvement, 24(5), pp. 4-16.
- [13] Montgomery, D.C., Runger, G.C., (1994). *Applied Statistics and Probability for Engineers*. Wiley, New York.
- [14] Montgomery, D.C., (1996), *Introduction to Statistical Quality Control* Third Edition, John Wiley and Sons, Inc.
- [15] Prof. Robert C. Leachman, (2005), *Statistical Process Control IEOR 130*, Methods of Manufacturing Improvement Spring,
- [16] Robert N. R. (1996). *Health Care Applications of Statistical Process Control: Examples Using the SAS System* SAS Institute Inc.Cary, North Carolina, USA.

- [17] Stephen P. D. (2005), *Statistical Process Control and Calcineurin Inhibitor Management*. The Journal of Pediatrics, Volume 146, Issue 6, June, Pages 720-721.
- [18] T.J. Harris, W.H. Ross, (1991). *The Canadian Journal of Chemical Engineering*. Statistical Process Control Procedures for correlated observations, 69 ,48–57.
- [19] T.J. Harris, W.H. Ross (1991), *Statistical process control procedures for correlated observations*, The Canadian Journal of Chemical Engineering 69, 48–57.
- [20] Noorlisa H. (2005), *Fault detection and Diagnosis via Improved Multivariate Statistical Process Control*. Methodology, Results and Discussions, 116-136.
- [21] S.W. Wang, F. Xiao (2004).” AHU Sensor Fault Diagnosis using Principal Component Analysis Method”, *Energy and Buildings* 36.147–160.
- [22] Stefan H., Steiner and R. J. M. (2000), *Monitoring Processes with Data Censored Due To Competing Risks Using EWMA Control Charts*. In industry, process monitoring is widely employed to rapidly. Sciences University of Waterloo Waterloo, N2L 3G1 Canada, March 30.
- [23] Yeung, K. Y (2001).”An Empirical Study of Principal Component Analysis for clustering gene expression data”.*Computer Science and Engineering*.University of Washington

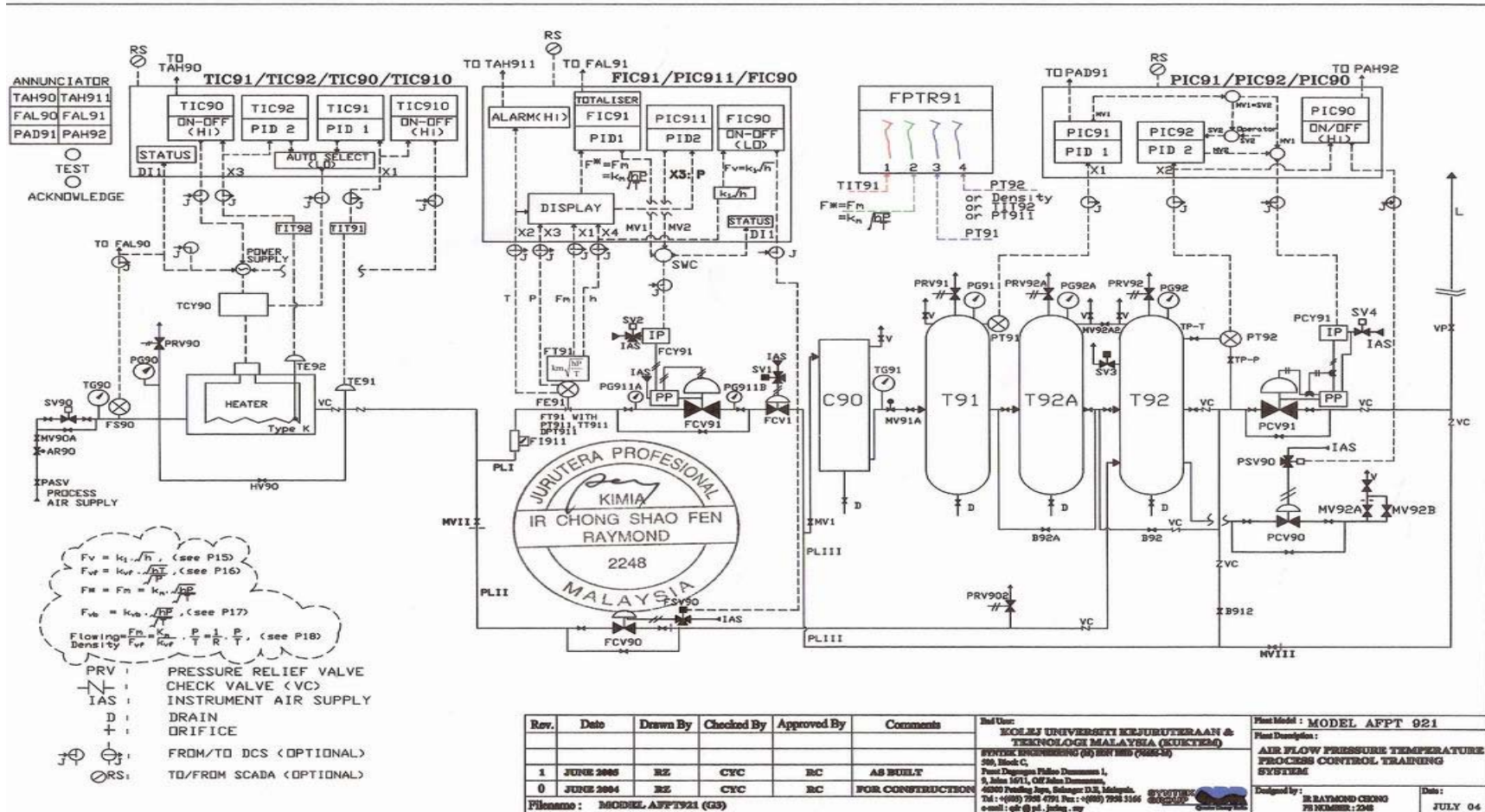
APPENDIX I

CASE STUDE OF AIR FLOW PRESSURE TEMPERATURE (AFPT921 PLANT)

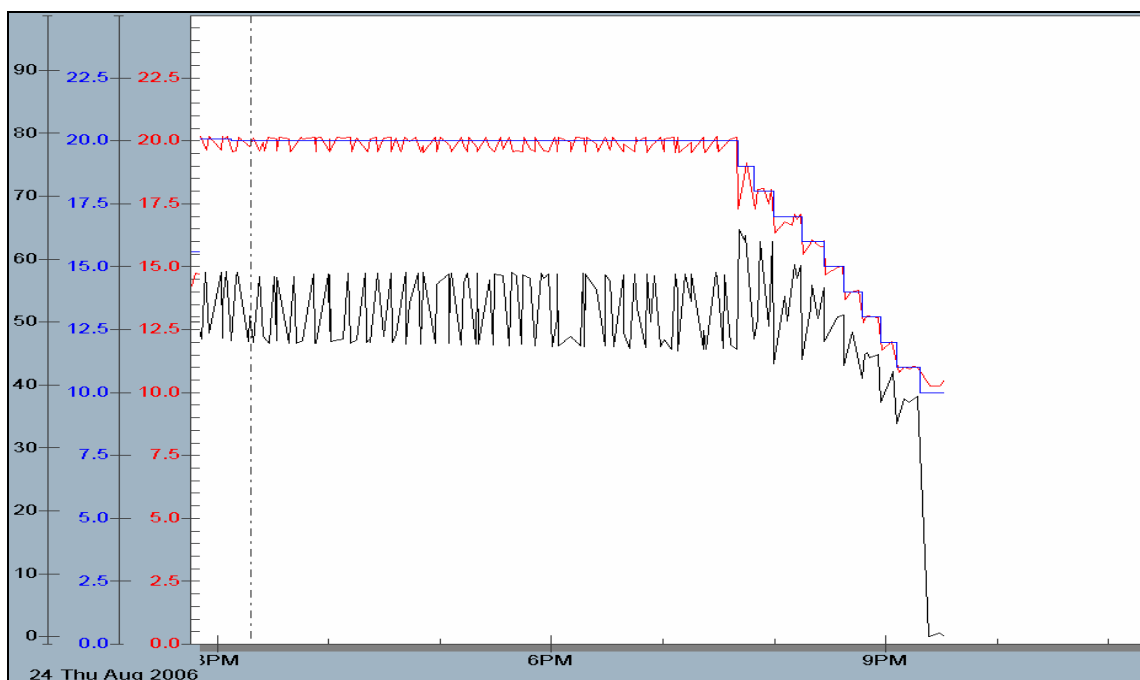
CASE STUDY OF AFPT 921 PLANT (PROCESS CONTROL TRAINING SYSTEM)



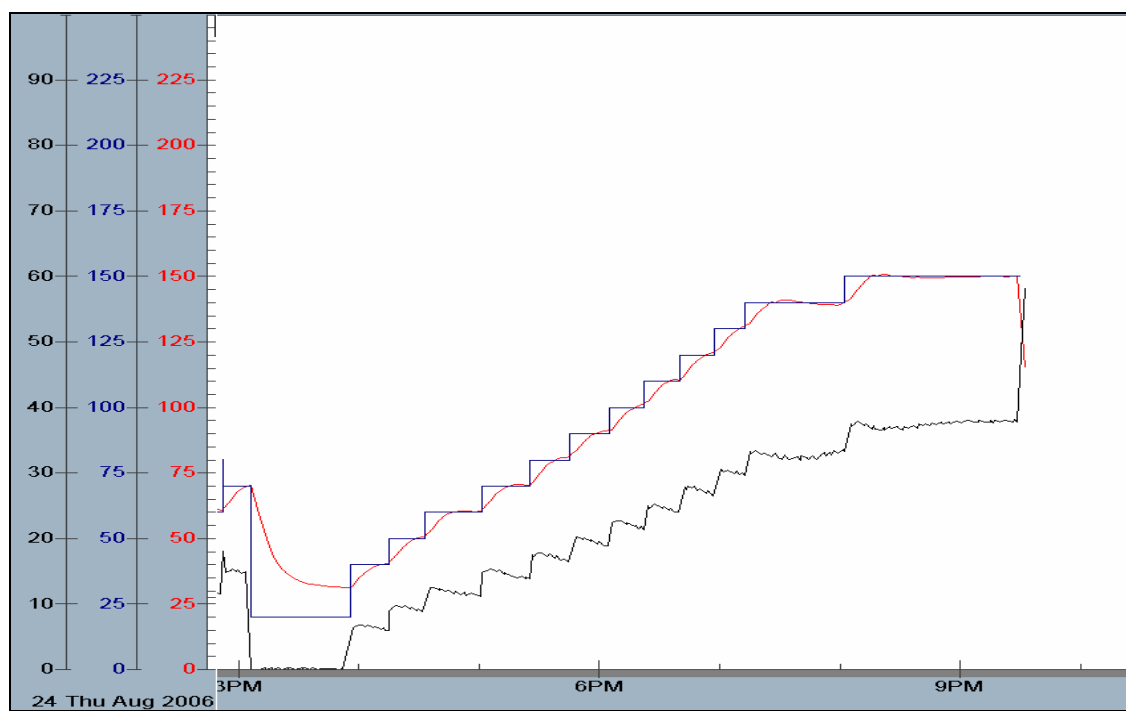
Process Instrumentation Design I



Process Instrumentation Design II



PIC92 Pressure vs Time



TIC92 Temperature vs Time

APPENDIX II

CODING FOR GAS DENSITY MEASUREMENT MODEL USING PCA TECHNIQUE IN MATLAB

CODING FOR PCA TECHNIQUE USING MATLAB

```
>>
```

```
x = temperature
```

```
Columns 1 through 13
```

```
326.0858 328.1410 330.1390 332.0512 334.0514 336.1364 338.0842 340.0977
342.1410 344.0711 346.1186 348.1401 350.1201
```

```
Columns 14 through 26
```

```
372.0821 354.0911 356.0714 358.1211 360.1012 362.1314 364.0509 366.1210
368.0844 370.1332 372.1245 374.0613 376.0841
```

```
Columns 27 through 39
```

```
378.0501 380.1493 382.1242 384.1150 386.1047 388.0710 390.0811 392.0944
394.1252 396.1201 398.1342 400.0821 402.0710
```

```
Columns 40 through 50
```

```
404.1267 406.0643 408.0712 410.1060 412.1030 414.1439 416.1485 418.1412
420.1201 422.0943 423.1672
```

```
y = pressure
```

```
1.0e+005 *
```


Columns 1 through 13

1.3728	1.2854	1.2862	1.3157	1.2794	1.3125	1.2345	1.2415	1.2450
1.2472	1.2190	1.1456	1.1631					

Columns 14 through 26

1.1737	1.1760	1.1515	1.0963	1.1333	1.1071	1.0992	1.1685	1.0505
1.0554	1.1329	1.1538	0.9433					

Columns 27 through 39

0.9643	0.9602	0.9652	0.9591	0.9616	0.8865	0.8952	0.8960	0.8985
0.8983	0.8189	0.8301	0.8296					

Columns 40 through 50

0.8280	0.8281	0.7450	0.7542	0.7571	0.7580	0.7598	0.7201	0.7132
0.7124	0.7113							

M = x

375.4849

M = y

1.0168e+005

z = covariance matrix

1.0e+008 *

0.0000 -0.0056

-0.0056 3.9684

ans =

1.0e+008 *

0.0000

3.968

x =

-1.0000 -0.0014

-0.0014 1.0000

y =

1.0e+008 *

0.0000 0

0 3.9684

PC = principal component

0.7061 -0.7081

0.7081 0.7061

PC = principal component

0.7061 -0.7081

0.7081 0.7061

SCORE =

-0.7071 0

0.7071 0

latent =

1.0000

0

tsquare =

1.0000

PC = principal component

-0.0000 1.0000

1.0000 0.0000

PC = principal component

-0.0000 1.0000

1.0000 0.0000

SCORE =

1.0e+008 *

-1.9842 0

1.9842 0

latent =

1.0e+016 *

7.8742

0

tsquare =

1.0000

pc = principal component

1 0

0 1

pc = principal component

1	0
0	1

latent =

1
1

explained =

50
50

pc =

0	1
1	0

pc =

0	1
1	0

latent =

1.0e+008 *

3.9684

0.0000

explained =

100.0000

0.0000

C = correlation coefficient

1.0000 -0.9788

-0.9788 1.0000

k1 = 1

k2 = 0.8171

P = 1.0e+005 *

Columns 1 through 13

-1.3404 -1.2549 -1.2556 -1.2845 -1.2489 -1.2813 -1.2050 -1.2117 -1.2152 -
1.2173 -1.1897 -1.1178 -1.1350

Columns 14 through 26

-1.1451 -1.1475 -1.1235 -1.0695 -1.1057 -1.0800 -1.0723 -1.1400 -1.0245 -
1.0293 -1.1051 -1.1256 -0.9195

Columns 27 through 39

-0.9401 -0.9360 -0.9409 -0.9349 -0.9374 -0.8639 -0.8723 -0.8731 -0.8755 -
0.8753 -0.7976 -0.8085 -0.8080

Columns 40 through 50

-0.8064 -0.8065 -0.7251 -0.7341 -0.7369 -0.7377 -0.7395 -0.7007 -0.6939 -
0.6931 -0.6920