Dynamic PTDF Implementation in the Market Model

Diyun Huang

(4033388)

Master Thesis Electrical Power Engineering

September 12, 2011

Delft University of Technology

Faculty of Electrical Engineering, Mathematics and Computer Science

Table of Contents

Chapter 1 Introduction 4
1.1 Background 4
1.1.1 Electrical Power System 4
1.1.2 Liberalization of the Electricity Sector and Cross-border Trade 5
1.1.3 Congestion Management and Flow Based Market Coupling 6
1.2 Problem Formulation and Objectives 7
1.3 Tools 10
1.4 Thesis Outline 10
Chapter 2 Power Transfer Distribution Factor (PTDF) and Literature Survey 13
2.1 Introduction 13
2.2 Nodal Network Model and Nodal PTDF Matrix 13
2.2.1 Nodal Transfer PTDF Matrix and Nodal PTDF Matrix 15
2.2.2 Derivation and Computation of Nodal PTDF Matrix 16
2.3 Zonal Network Model and Zonal PTDF Matrix 19
2.3.1 Zonal Network Models 19
2.3.2 Zonal PTDF Matrix 22
2.4 Security and Safe PTDF 23
2.5 Zonal PTDF Study Cases in Powrsym4 23
2.5.1 PTDF without Transmission Capacity Limit 23
2.5.2 PTDF with Transmission Capacity Limit 25
2.6 Conclusion 26
Chapter 3 Calculation Methods for Zonal PTDF 28
3.1 Introduction 28
3.2 New England Model Description 28
3.3 Influencing Factors of Zonal PTDF 30
3.3.1 Seasonal Variation 30
3.3.2 Day/Night Variation 32
3.4 Classic PTDF Calculating Method 33
3.4.1 Generation Shift Method 35
3.4.2 Implementation 36
3.5 GSK PTDF Calculation Method 39
3.6 Conclusion 42
Chapter 4 Data Analysis Methodology and Dynamic PTDF Family Identification 44
4.1 Introduction 44
4.2 Input Data Preparation for Data Analysis 45
4.2.1 Hour Number Segmentation 45
4.2.2 Selection of Parameters for Data Analysis Using Linear Regression 46
4.3 Typical Hours Selection Process and Study Case 47
4.4 Conclusions 59
Chapter 5 Economic Dispatch with Dynamic PTDF Logic 63
5.1 Introduction 63
5.2 Economic Dispatch Model with classic PTDF in Matlab 63
5.2.1 Linear Programming 63
5.3 Implementation of Dynamic PTDF Matrix on UC-ED Decision Making 66

5.4 Conclusion.. 73
Chapter 6 Conclusions and Recommendations 76
Chapter 7 References 78
Mathematical Methods 80
Linear Regression 80
Partitioning Methods 84
Optimal Number of Clusters 87

Chapter 1 Introduction

1.1 Background

1.1.1 Electrical Power System

The vertically structured electric power system consists of four principal parts: generation, transmission, distribution and load.

Generation is the part where the primary energy source is utilized to convert mechanical, chemical or other form of energy into the electric energy. Another way of description of generation is the large-scale electricity production in a central plant. A power plant has one or more units, with an individual turbine generator in each unit to use steam, heated gas, falling water or wind to generate electric power [1].

Transmission is the process of transporting electricity in bulk from the generation plants to supply other principal parts. The transmission network has high voltage ratings, connects the bulky central power plants with transmission substations over long distance and supplies the distribution subsystems. The transmission network voltage rating is different in different countries. In the Netherlands, the transmission network voltage is 110 KV and higher [9]. As the Dutch transmission system operator and independent Dutch transmission grid administrator, Tennet interconnects all the regional networks and European grid [1].

Distribution is the act of delivering electric energy from substations in transmission system to consumers. It always has the medium or low voltage ratings over short distance. A load transforms the electrical energy into other forms of energy, such as mechanical energy, light, heat or chemical energy. As grid users, individual loads in the system are clustered to three types: residential loads, commercial and industrial loads and electric railways [1].

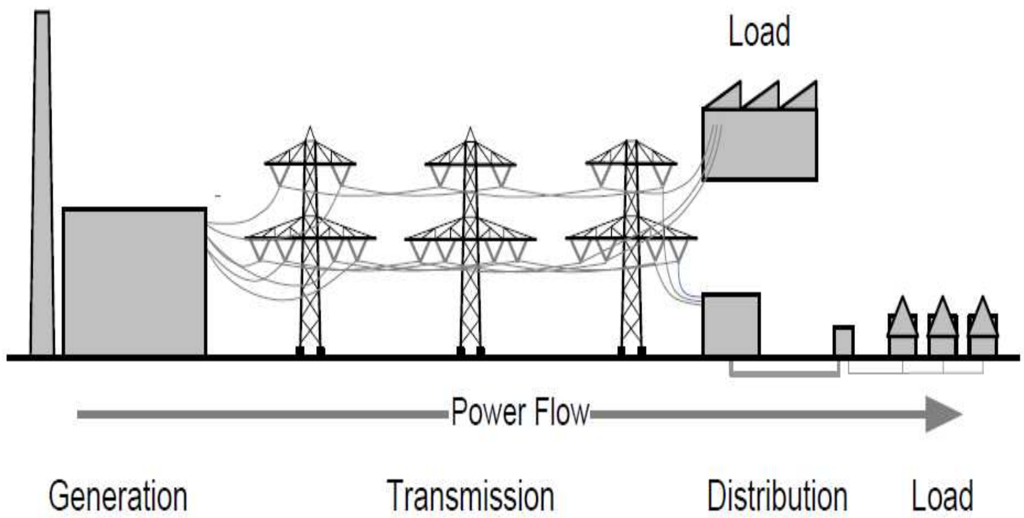

Figure1.1 Electrical Power System [2]

1.1.2 Liberalization of the Electricity Sector and Cross-border Trade

In the past decades, European countries have gone through a long way in liberalization of the electricity market [3]. A major change is the unbundling of the vertically integrated system into four separated segments. Generation ownership is separated from the transmission network, so that competition could be introduced in the supply side.

Besides this, the European Union is actively promoting the process of achieving an integrated electricity market [4]. The integrated electricity market is seen as a means to bring down the electricity price for the well being of European citizens and to tackle climate change while maintaining energy security.

To achieve an integrated electricity market, regional market coupling is the most realistic method in short and medium term as suggested by ETSO and EuroPEX [5]. Therefore, the interconnections between countries become the backbone to couple regional market and enhance competition between market players in the member states as it facilitates the electricity export from low-price market area to a high-price market area. Regulation

1228/2003 is focused on cross border trade of electricity and states the principles of cross-border trade [6].

1.1.3 Congestion Management and Flow Based Market Coupling

However, the tie lines were initially built for system security, in particular to share the reserves required by safety standards, instead of facilitating large scale inter-regional trade [8]. The European Commission has identified that insufficient interconnection capacity linking member states is hampering the formation of integrated European electricity market (Commission of the European Communities 2003).

Congestion arises when the flow on certain interconnection reaches its capacity limit with increasing cross-border trade. Yet it is infeasible to build infinite interconnection capacities because of economic constraints [7]. In addition, due to the physical characteristic of electricity, it does not simply follow contract path (from the sending region to the withdrawal region directly), which brings more complexity to the congestion management of electricity network compared with congestion in other infrastructure.

To handle the congestion management, ETSO and EuroPEX proposed the flow based market coupling, in which the regional markets are coupled and TSOs need to take into account the physical power flows on the borders of different markets [5]. In a single-price region, which always conforms to the boundary of a country or transmission system operator, the nodes are aggregated and represented by one equivalent node. Then the interconnections between a pair of regions are substituted by a virtual link between the two regional nodes. The limits are put on the aggregated virtual inter-regional links to represent transmission constraints of cross-border lines. Zonal PTDF (Power Transfer Distribution Factor) is used to link the commercial transaction and the resulted power flows on cross-border links. In short, zonal PTDF represents the flow property of the flow-based model. With the simplified transmission model and its ability to approximate inter regional flow, market coupling enables trade between regional markets to achieve the maximum economic efficiency.

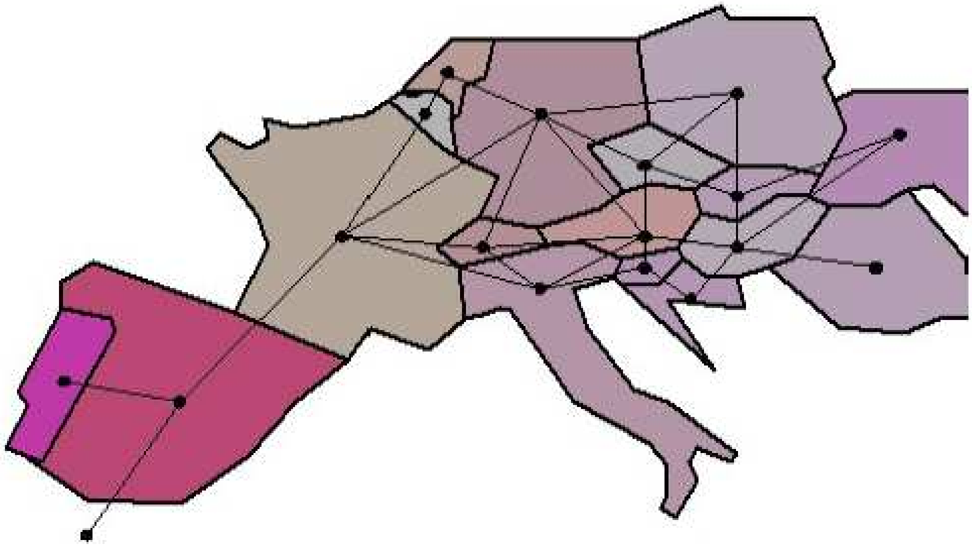

Figure1.2 Flow based model [9]

Flow based market [5] has the following advantages. First of all, flow based model maximize the inter-regional capacity allocation for network users without undermining the system security. At the same time, market coupling offers efficiency advantage as it provides competition between regions.

1.2 Problem Formulation and Objectives

Nowadays, the electricity market models such as Powrsym3 do not reflect the physical properties and limitations of the network. The market model assumes that the power flow follows contract path instead of respecting the Kirchhoff's law. Thus estimated inter regional power flow in the market model is largely deviated from the real flow. This drawback may lead to biased capacity allocation for cross-border trade which sacrifices the economic efficiency or causes severe condition for the network.

As an important feature in the flow based market coupling to solve the above mentioned problem, PTDF (Power Transfer Distribution Factor) translates the cross-border (or inter-regional) transaction into the power flow distribution in the cross-border links (See Congestion management and Flow based market coupling**).** It gives the fraction of the amount of transaction from one region to another that flows over a given inter-regional link and represents the sensitivity of the operational transmission grid status towards a system change of injection and withdrawal of power. Adopting the PTDF in the market model will be lead to more accurate network representation in the market model.

The impetus of this master thesis lies in the fact that using only one fixed PTDF to estimate power flow after the unit commitment and economic dispatch (UC-ED) is run in the market model will still bring a large error. In the study case, there is an average error of 16.9% for power flow estimation. 11.6% of the hours have an error of more than 100MW in the time series with an average real power flow of 524MW as shown in Figure 1.3.

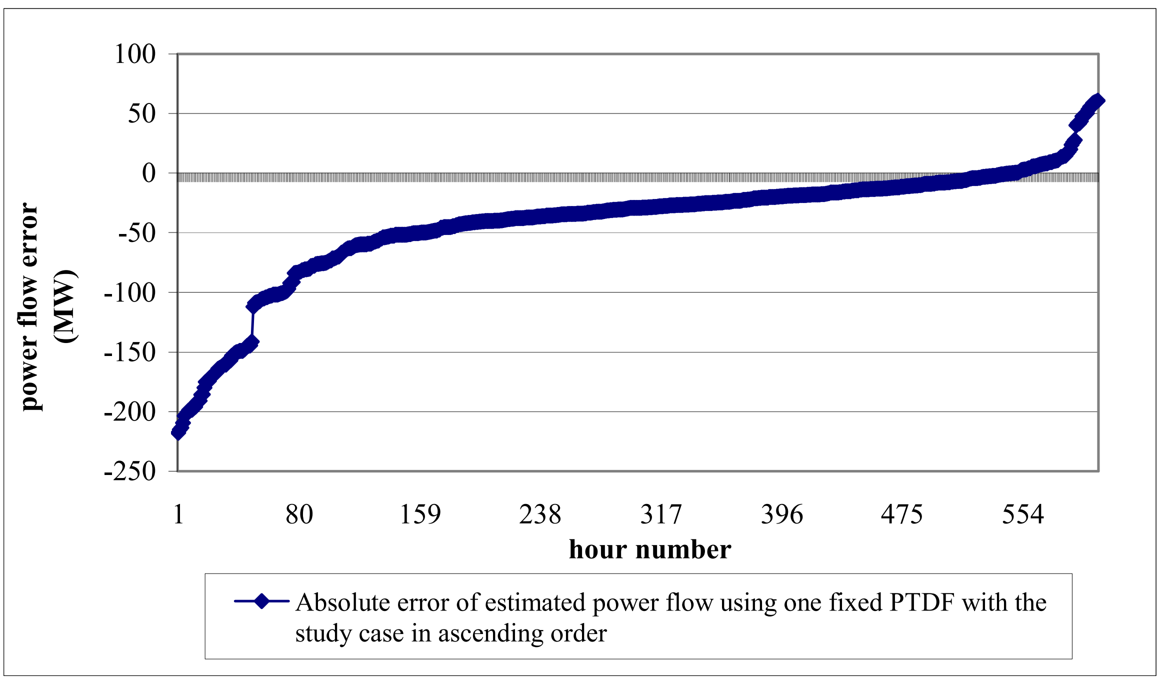

Figure1.3 Absolute error of estimated power flow using one fixed PTDF with the study case in ascending order

Therefore, the concept of dynamic PTDF family is developed in this master thesis to create a group of typical PTDF matrices which can be chosen dynamically to represent random hours in market simulation.

The objective of the master thesis is *to create a family of PTDF matrices which can be dynamically chosen during the market simulation in order to have a more accurate representation of the grid.*

Several steps are taken to achieve the objective:

1. Create dynamic PTDF family
    - Find the methodology to look for the most typical scenarios with input data from market model

    - Calculate the PTDF values in these typical scenarios and form the dynamic PTDF family

    - Define the rule to choose the matrix from dynamic PTDF family that approximates real hourly PTDF value in a weekly and yearly time scope

2. Validation of simulation result improvement using dynamic PTDF family
    - Build economic dispatch and power flow calculation model of New England case with PTDF in Matlab

    - Use the hourly PTDF chosen from dynamic PTDF family in the Matlab model to calculate the economic dispatch and power flow on inter-area links

    - Compare it with the PowrSym4 result without implementing PTDF, with the simple PTDF logic of using only one PTDF matrix and the real power flow calculated by PSS/E and see the improvement of accuracy with dynamic PTDF family.

1.3 Tools

The main tools used to carry out the research for this master thesis are mainly Tennet's market modeling software PowrSym4, transmission system modeling software PSS/E, Matlab.

PowrSym4 is a multi-area, multi-fuel production cost simulation model co-developed by Operation Simulation Associates, Inc. of the USA and Tennet. It provides the simulation accuracy and level of detail suitable for both short term operation studies and long term planning. There are mainly three time horizons for simulation in PowrSym4: annually, weekly and hourly. The result of annually simulation is applied for reliability calculation and maintenance scheduling. The weekly result is for outage simulation, hydro scheduling and energy storage units, and the hourly horizon is used for simulation of unit commitment and economic dispatch [10].

PSS/E is the software tool for power system analysis widely used by transmission planning and operation engineers. The PSS/E power flow package provides modeling in a user friendly and convenient environment. Moreover, PSS/E 31 includes a Python API, which enables the automation of PSS/E from Python code.

Matlab is a technical computing language for numeric computation, data analysis and data visualization. The toolboxes of Matlab enable the users to solve problems in particular application areas by category.

1.4 Thesis Outline

The content of the master thesis is divided into six chapters. Each chapter starts with a short introduction to present the core topics and emphasize the contribution of this master thesis.

The first chapter is the general introduction. At the beginning, it gives background information on electricity market liberalization and flow-based market coupling. Hence we introduce the general concept of PTDF matrix and reveal the importance of PTDF matrix as a useful instrument in congestion management. Later the objective of the thesis is outlined and different kinds of software used in this master thesis are briefly introduced.

Chapter two presents the state of art on PTDF. Several types of PTDF and their corresponding network representations from literatures are introduced and the relationship between them is deduced.

In Chapter 3, two different zonal PTDF calculation methods are investigated. The influence factors of zonal PTDF are discussed. Then the classic PTDF calculation method and GSK PTDF calculation method are presented, followed by the description of network model used in this master thesis. In this part, different physical implication of the two zonal PTDF methods as well as difference in application is emphasized. Finally, weekly power flow estimation result by the two methods is shown.

In Chapter 4, the methodology is developed to create a dynamic PTDF family; a rule is defined choose a representative PTDF matrix from the family for a random hour. The effectiveness of the dynamic PTDF family selected is shown in the study case by comparing the estimated power flow using the dynamic PTDF matrices, a fixed PTDF matrix and the real power flow calculated in PSSE, given the economic dispatch run by Powrsym4 simulation without PTDF logic. Significant reduction of power flow estimation error can be seen when adopting dynamic PTDF family compared with using a fixed PTDF matrix or without PTDF matrix.

In Chapter 5, an economic dispatch model is built in Matlab with PTDF logic and the algorithm is presented. Then the dynamic PTDF logic is implemented in the economic dispatch model is and its influence on the UC-ED is investigated. A comparison is made between adopting the typical scenario from dynamic PTDF family and the scenario of the first hour of the study year in the economic dispatch model. In the end, an implementation scheme of the dynamic PTDF family for the UC-ED decision making is proposed.

In chapter 6, summary and conclusion from this master thesis are presented as well as recommendations for future research work.

Chapter 2 Power Transfer Distribution Factor (PTDF) and Literature Survey

2.1 Introduction

In this chapter, a classification of network representations and their corresponding PTDF matrices are discussed. The chapter is organized as follows. First, the nodal PTDF is introduced, and the physical meaning as well as certain characteristics of the nodal PTDF matrix is explained with a simple three node network, followed by calculating the transfer nodal PTDF matrix and its relationship with the nodal PTDF matrix. The nodal PTDF matrix is derived from the load flow equations and the computation method is presented. Then, two zonal network models and corresponding forms of PTDF are discussed. For security issues, the concept of 'safe' PTDF is introduced. In the last part, two examples designed initially to test the Powrsym4 software are presented for illustrating how PTDF logic is implemented in the market model.

2.2 Nodal Network Model and Nodal PTDF Matrix

The nodal PTDF matrix states the influence of each nodal power injection on a given individual line [11]. There is a reference node, with which all power transactions are made between each injection node. By assuming a reference node, the nodal PTDFs are limited to only nodal injections (withdrawn is always in reference node) instead of all combinations of transactions between each pair of nodes. As will be derived in section 2.2.2, the nodal PTDF values only depend on the network topology and branch parameters.

$$\vec{P}_f = PTDF * \vec{P}_{inj} = PTDF * (\vec{P}_{generation} - \vec{P}_{load}) \tag{2.1}$$

Where

$\vec{P_f}$ is the vector of power flows.

$\vec{P_{inj}}$ is the vector containing the power injection into the nodes.

$\vec{P_{generation}}$ is the generation vector.

$\vec{P_{load}}$ is the load vector.

For a system with N buses and M branches, the dimension of the nodal PTDF matrix is $M \times N$. In a 3 node system as in Figure 2.1, if we assume that all lines are identical and we take node 2 as reference node, the nodal PTDF matrix is shown in Table 2.1.

	Node 1	**Node 2**	**Node 3**
Line 1-2	2/3	0	1/3
Line 1-3	1/3	0	-1/3
Line 2-3	-1/3	0	-2/3

Table 2.1 Nodal PTDF matrix of the three node network

The first column reveals the power flow distribution when 1 MW is sent from node 1 to the reference node (node 2).The negative sign corresponding to line 2-3 shows that the power flow direction is from node 3 to node 2, opposite to the predefined direction.

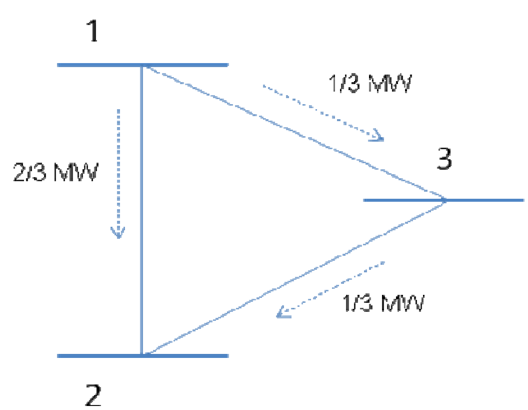

Figure 2.1 Three node network

The following equations need to be satisfied in column j when a transfer is made from the sending node j to the reference node.

$$\sum_{all\ lines\ from\ sending\ node} PTDF(sending\ node) = 1(100\%)$$
$$\sum_{all\ lines\ from\ reference\ node} PTDF(reference\ node) = -1(-100\%) \tag{2.2}$$

When we look at the nodal PTDF in Table 2.1, in the first column, node 1 is the sending node. The sum of PTDF values of the lines from node 1(line 1-2, line1-3) is 1. The PTDF of line 2-1 equals the negative PTDF value on the line 1-2. So the sum of the PTDF values of the lines from the reference node (node 2) is -1.

2.2.1 Nodal Transfer PTDF Matrix and Nodal PTDF Matrix

The transfer Nodal PTDF matrix can be derived from a nodal PTDF matrix easily. Power flow distributions of all combinations of transactions between each pair of nodes are taken into account in the transfer nodal PTDF matrix. The influence of transfer between node i and node j on line n-m between node n and node m can be derived by equation 2.3 [11].

$$PTDF_{i-j,line\ n-m} = PTDF_{i,\ line\ n-m} - PTDF_{j,line\ n-m} \tag{2.3}$$

Where

$PTDF_{i-j,line\ n-m}$ is the nodal transfer PTDF from node i to node j on the line n-m.

$PTDF_{i,\ line\ n-m}$ is the nodal PTDF of node i on the line n-m.

The dimension of the transfer nodal PTDF matrix in a system with N bus and M lines is $(N*(N-1))\times M$.

Though the size of the nodal transfer matrix is largely increased compared with the nodal PTDF matrix, it is easier to investigate the effect of changing load demand on line power flow and to optimize the system cost. For example, for the 3 node system with equal line impedance in Figure 2.1, the transfer nodal PTDF matrix T is such as in Table 2.2.

	Transfer from 1 to 2	Transfer from 1 to 3	Transfer from 2 to 3	Transfer from 2 to 1	Transfer from 3 to 1	Transfer from 3 to 2
Line 1-2	2/3	1/3	-1/3	-2/3	-1/3	1/3
Line 1-3	1/3	2/3	1/3	-1/3	-2/3	-1/3
Line 2-3	-1/3	1/3	2/3	1/3	-1/3	-2/3

Table 2.2 Nodal Transfer PTDF matrix of the three node network

Assume that the line from node 1 to node 3 is at the point that is almost or already congested. Node 1 has the generation with the lowest production cost. Node 2 has the most expensive generation and thus is an importer. We can see directly from the transfer Nodal PTDF matrix that transfer from node 1 to node 2 would further burden the congested line, but a transfer from node 3 to node 2 would alleviate the stress.

We can make a transformation matrix L to link the nodal PTDF in table 2.1 and the nodal transfer PTDF in Table 2.2.

	Transfer from 1 to 2	Transfer from 1 to 3	Transfer from 2 to 3	Transfer from 2 to 1	Transfer from 3 to 1	Transfer from 3 to 2
Node 1	1	1	0	-1	-1	0
Node 2	-1	0	1	1	0	-1
Node 3	0	-1	-1	0	1	1

Table 2.3 Transfer matrix L of three node network

Apparently, we can obtain the transfer Nodal PTDF by:

$$PTDF_{nodal}^{transfer} = PTDF_{nodal} * L \tag{2.4}$$

2.2.2 Derivation and Computation of Nodal PTDF Matrix

2.2.2.1 Derivation of Nodal PTDF Matrix

The power flow equations are based on Ref. [12]-[14]

The typical element Y_{km} in the $N \times N$ bus admittance matrix is:

$$Y_{km} = |Y_{km}| \angle \theta_{km} = |Y_{km}| \cos \theta_{km} + j |Y_{km}| \sin \theta_{km} = g_{km} + jb_{km} \tag{2.5}$$

The voltage at bus k of the system in polar coordinates is given by

$$V_k = \left|V_k\right| \angle \delta_k = \left|V_k\right| (\cos\delta_k + j\sin\delta_k) \qquad (2.6)$$

The real power flow entering the network at typical bus k is:

$$P_k = \sum_{m=1}^{N} \left|Y_{km} V_k V_m\right| \cos(\theta_{km} + \delta_m - \delta_k) \qquad (2.7)$$

Derivation of the net real power entering the bus k yields:

$$\Delta P_k = 0 = V_k^2 g_{kk} + V_k \sum_{\substack{m=1 \\ m\neq k}}^{N} \left(V_m \left[g_{km}\cos(\delta_k - \delta_m) + b_{km}\sin(\delta_k - \delta_m)\right]\right) - P_{Gk} + P_{Lk} \qquad (2.8)$$

A PTDF is the derivation of branch flow with respect to a unit change in nodal injection.

$$PTDF_{km} = \Delta P_{km} = \left[\frac{\partial P_{km}}{\partial V_k}\right]\Delta V_K + \left[\frac{\partial P_{km}}{\partial V_m}\right]\Delta V_m + \left[\frac{\partial P_{km}}{\partial \delta_k}\right]\Delta \delta_K + \left[\frac{\partial P_{km}}{\partial \delta_m}\right]\Delta \delta_m \qquad (2.9)$$

The derivation of the power flow between nodes k and m yields:

$$\left[\frac{\partial P_{km}}{\partial \delta_m}\right] = V_k V_m \left[g_{km}\sin\left(\delta_k - \delta_m\right) - b_{km}\cos\left(\delta_k - \delta_m\right)\right] \qquad (2.10)$$

$$\left[\frac{\partial P_{km}}{\partial \delta_k}\right] = V_k V_m \left[-g_{km}\sin\left(\delta_k - \delta_m\right) + b_{km}\cos\left(\delta_k - \delta_m\right)\right] \qquad (2.11)$$

$$\left[\frac{\partial P_{km}}{\partial V_m}\right] = V_k \left[g_{km}\cos\left(\delta_k - \delta_m\right) + b_{km}\sin\left(\delta_k - \delta_m\right)\right] \qquad (2.12)$$

$$\left[\frac{\partial P_{km}}{\partial V_k}\right] = 2V_k g_{kk} + V_m \left[g_{km}\cos\left(\delta_k - \delta_m\right) + b_{km}\sin\left(\delta_k - \delta_m\right)\right] \qquad (2.13)$$

Decoupled load flow assumes the following [1]:

- the line conductances are much smaller than the line susceptances : $g_{km}\sin(\delta_k - \delta_m) \ll b_{km}\cos(\delta_k - \delta_m)$

- Differences between the voltage angles are small:

  $\sin(\delta_k - \delta_m) = \delta_k - \delta_m$ and $\cos(\delta_k - \delta_m) = 1$

The approximations of decoupled load flow give the following result:

$$\left[\frac{\partial P_{km}}{\partial \delta_m}\right] = -b_{km}$$

$$\left[\frac{\partial P_{km}}{\partial \delta_k}\right] = b_{km} \tag{2.14}$$

$$\left[\frac{\partial P_{km}}{\partial V_k}\right] = 0$$

$$\left[\frac{\partial P_{km}}{\partial V_m}\right] = 0$$

Thus PTDF can be expressed as:

$$PTDF_{km} = b_{km} * (\Delta\delta_k - \Delta\delta_m) \tag{2.15}$$

A linear relationship can be written between the bus injection and the bus angle:

$$\delta = Y^{-1} * P_B \tag{2.16}$$

$$\Delta\delta = Y^{-1} * \Delta P_B \tag{2.17}$$

$$PTDF_{km} = b_{km} * Y^{-1}\Delta P_B \tag{2.18}$$

Assume bus n is the slack bus and ΔP_B=1, and c_{ij} is the element (i,j) of the augmented matrix of Y^{-1} where a dummy row and column are added at the slack bus (Y^{-1} is not invertable), so we can have the PTDF value on the branch that links node k and node m by equation 2.19:

$$PTDF_{km} = \sum_{i=1}^{n-1} b_{km}(c_{ki} - c_{mi}) \tag{2.19}$$

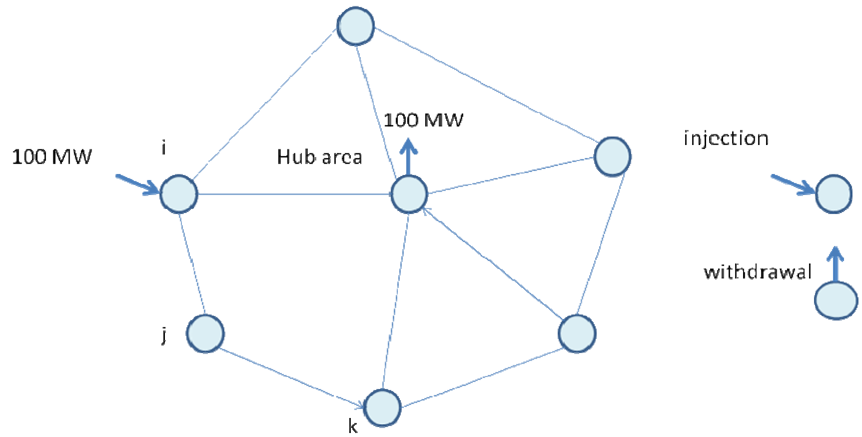

Figure 2.2 Nodal network representation with injection and withdrawal

The nodal PTDF value that represents the influence of injection at node i on line j-k can be calculated using the following equation:

$$PTDF_{i,line\ j-k} = \frac{PF'_{j\rightarrow k} - PF^{\mathrm{Re}f}_{j\rightarrow k}}{100} \tag{2.20}$$

$PF'_{j\rightarrow k}$ Flow on the line $j \rightarrow k$ after the injection of 100 MW in node i and withdrawal at node j [MW]

$PF^{\mathrm{Re}f}_{j\rightarrow k}$ Flow on the line $j \rightarrow k$ in reference case

2.3 Zonal Network Model and Zonal PTDF Matrix

2.3.1 Zonal Network Models

A nodal network is depicted in Figure 2.3. The network is lossless and comprises five nodes and seven lines. Node 3 is the reference node.

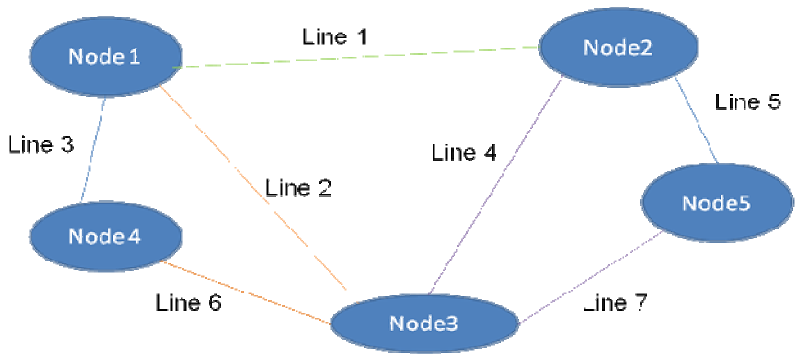

Figure 2.3 Nodal representation of network

	Node 1	Node 2	Node 3	Node 4	Node 5
Line 1	0.285714	-0.28571	0	0.142857	-0.14286
Line 2	0.47619	0.190476	0	0.238095	0.095238
Line 3	0.238095	0.095238	0	-0.38095	0.047619
Line 4	0.190476	0.47619	0	0.095238	0.238095
Line 5	0.095238	0.238095	0	0.047619	-0.38095
Line 6	-0.2381	-0.09524	0	-0.61905	-0.04762
Line 7	-0.09524	-0.2381	0	-0.04762	-0.61905

Table 2.4 Nodal PTDF matrix of the five node network

The zone to line representation of the grid aggregates the nodes into zones and the real interconnection lines between nodes in different zones. It is named Flow based market coupling (FBMC) representation of the grid by Ehrenmann and Smeers [7]. The network in Figure 2.4 shows the zone-to-line representation of the grid.

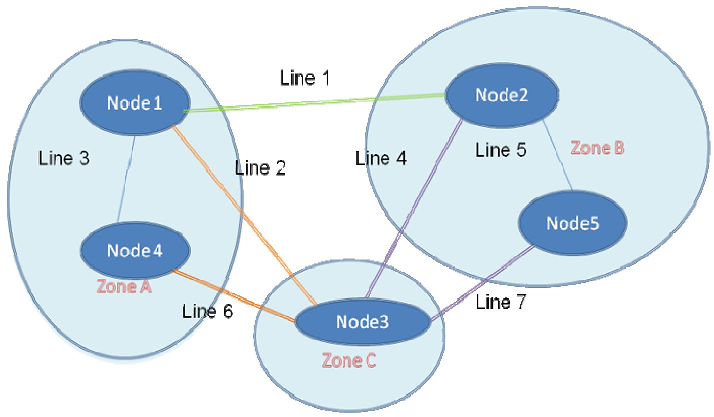

Figure 2.4 zone-to-line representation of the grid

In zonal representation of the grid such as given in Figure 2.5, nodes are aggregated into zone and the different lines linking the nodes in two zones are further aggregated into a single inter-zonal link.

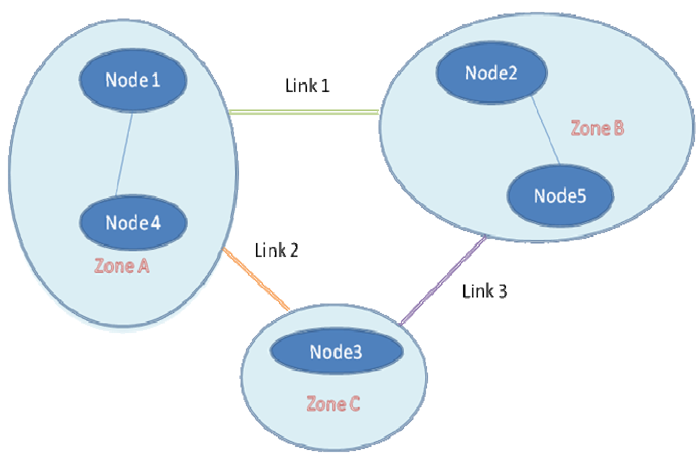

Figure 2.5 Zonal representation of the network

There are some simplifications made for the zonal model that we should pay attention to:

- Zonal representation of grid assumes homogeneous generation and load in a zone in the sense that the same amount of power injection or withdrawal in any nodes within a zone is considered to have the same influence on the inter-zonal links. It is not the reality and can bring substantial inaccuracy in large geographical zone.
- Internal congestion in a zone is neglected, and since the focus is on inter-zonal trade, injections and withdrawals between the nodes of a zone are seen as transfers between representative zonal node and be paid a zonal price
- One inter-zonal link is used to represent interconnections between a pair of zones,

thus violation of individual line capacity is not considered.

2.3.2 Zonal PTDF Matrix

2.3.2.2 Zone-to-line PTDF Matrix

Corresponding to the grid in Figure 2.4, assume that zone C is the hub zone. If each node in a zone is considered to contribute with the same amount of power to the inter-zonal transfer, the zone-to-line PTDF has the form shown in Table 2.5.

	Zone A	Zone B	Zone C
Line 1	0.2143	-0.2143	0
Line 2	0.3571	0.1429	0
Line 4	0.1429	0.3571	0
Line 6	-0.4286	-0.0714	0
Line 7	-0.0714	-0.4286	0

Table 2.5 Zone-to-line PTDF matrix of the network in Figure 2.4

2.3.2.3 Zonal PTDF Matrix

A zonal PTDF is the derivative of flow on the inter-zonal link with respect to a unit change in injection in the each zone (withdrawal always takes place in the hub zone).

The zonal PTDF matrix varies with topology, operating points and zone building [15]. These influence factors of zonal PTDF will be discussed in detail in chapter 3.

Zonal PTDF are classified in a matrix similar to Nodal PTDF matrix where columns are indexed by zones and the rows represent inter-zonal links. For example, the zonal PTDF matrix for the network in Fig 2.5 has the following form:

	Zone A	Zone B	Zone C
Link 1	0.2143	-0.2143	0
Link 2	0.7857	0.2143	0
Link 3	0.2143	0.7857	0

Table 2.6 Zonal PTDF matrix example

The advantage of adopting zonal network model and using zonal PTDF is very obvious. First, the computation complexity is reduced and it does not require detailed network data such as line inductance. More importantly, by aggregating nodes into zones, the number of submarkets is decreased and trading is simplified [7].

But the zonal network model combined with the zonal PTDF matrix will cause a loss of accuracy compared with nodal representation of the grid. Moreover, the flows on inter-zonal links due to intra-zonal transactions are neglected.

2.4 Security and Safe PTDF

Ehrenmann und Smeers [7] introduced the concept of 'safe PTDF', in the belief that by using conservative PTDF values the flows can be always made feasible in the congested lines. The maximal nodal PTDF is used for exporting zone and minimal nodal PTDF is used for importing zone. But application of this 'safe PTDF' can lead to severe loss of economic efficiency as it overestimates the power flow on congested lines.

Duthaler [15] examined the flow overestimation as a result of applying 'safe PTDF'. First 6 snapshots from 2004 to 2007 were run and six PTDF were derived. Then the maximal PTDF value as from each element of the six matrices was derived by the average value. Table 2.7 shows the the overestimation effect in UCTE network:

	Transactions				
Flowgates	O->I	S->I	D->I	F->I	L->I
CH->IT	111%	112%	111%	111%	118%
CH->DE	112%	136%	113%	119%	119%
CH->FR	193%	118%	168%	118%	216%
CH->AT	128%	223%	128%	130%	124%
DE->FR	128%	175%	138%	126%	126%
DE->AT	112%	143%	125%	130%	119%
IT->FR	121%	122%	117%	108%	115%
IT->SI	111%	136%	121%	133%	110%
AT->SI	119%	145%	143%	145%	106%
	Central South				

Table 2.7 Overestimation by safe PTDF [15]

2.5 Zonal PTDF Study Cases in Powrsym4

2.5.1 PTDF without Transmission Capacity Limit

In a three zone lossless network with the generation capacity, generation cost and load such as in Table 2.9, generators are located in zone 1 and zone 2 and load is located at zone 3. Both zone 1 and zone 2 can be exporter with available generation capacity to supply the load in zone 3. Without any transmission capacity limit in the network, only the lowest cost zone (zone1) exports power to the load zone (zone 3), since the available generation capacity in zone 1 is enough to supply the total load. The unit commitment and economic dispatch output when we apply PTDF in lossless network in Powrsym4 is the same as the simulation result without implementing PTDF logic.

	Maximum Generation Level (MW)	Cost (Euro/MWh)	Load (MW)	Production (MW)
Zone 1	200	10	0	90
Zone 2	100	20	0	0
Zone 3	0	-	90	0

Table 2.8 Generation and load for three zones

Assume the system has the zonal PTDF matrix shown in Table 2.9 with zone 2 chosen to be the hub zone.

	Zone 1	Zone 2	Zone 3
Link 1	2/3	0	1/3
Link 2	1/3	0	-1/3
Link 3	-1/3	0	-2/3

Table 2.9 Zonal PTDF of the system

The inter-zonal power flows can be calculated by using equation 2.1 from section 2.2:

$$\begin{bmatrix} P_{f1} \\ P_{f2} \\ P_{f3} \end{bmatrix} = \begin{bmatrix} 2/3 & 0 & 1/3 \\ 1/3 & 0 & -1/3 \\ -1/3 & 0 & -2/3 \end{bmatrix} * \begin{bmatrix} 90 \\ 0 \\ -90 \end{bmatrix} = \begin{bmatrix} 30 \\ 60 \\ 30 \end{bmatrix} \tag{2.22}$$

$\vec{P}_{inj}$ refers to zonal injection. Positive sign refers generation and negative sign refers to load.

The inter-zonal power flow resulted from the transaction between zone 1 and zone 3 can be seen in Figure 2.6:

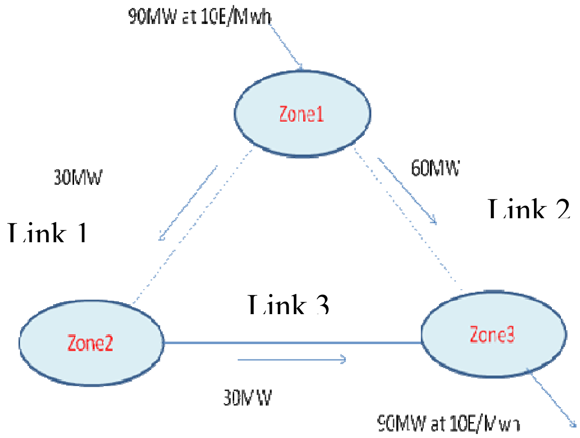

Figure 2.6 Inter-zonal power flows in the three zone network without transmission capacity limit

2.5.2 PTDF with Transmission Capacity Limit

If we consider the same network from section 2.8.1, if there is a 10 MW transmission capacity limit on the link between zone 1 and zone 2 as can be seen in Figure 2.7, the solution from section 2.8.1 with 90 MW power transfer from zone 1 to zone 3 will not be feasible anymore. But we can easily observe that a counter flow resulted from transaction between zone 2 and zone 3 can relieve the congestion on link 1.

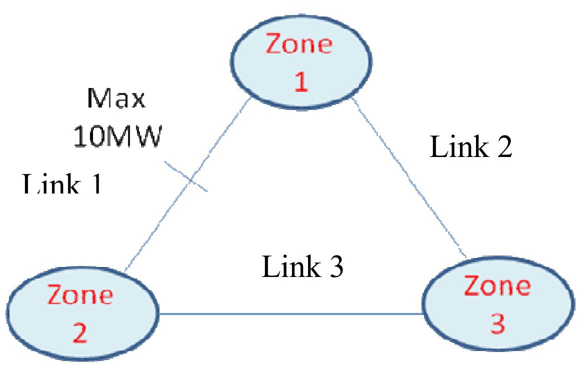

Figure 2.7 Three zone network with transmission capacity limit on one link

The optimal solution in this case will minimize the cost when trying to serve the system load without exceeding the transmission capacity limit. The zones with more expensive generation will start to produce, so the final unit commitment and economic dispatch result when using PTDF logic (as shown in Table 2.10) will be different from the simulation output without PTDF logic (as shown in Table 2.8).

	Maximum Generation Level (MW)	Cost (Euro/MWh)	Load (MW)	Production (MW)
Zone 1	200	10	0	60
Zone 2	100	20	0	30
Zone 3	0	-	90	0

Table 2.10 Generation and load for 3 zones with transmission capacity limit on one link and with PTDF logic

The inter-zonal power flows as a result of the power transfer from zone 1 to zone 3 and from zone 2 to zone 3 are illustrated respectively in Figure 2.8, along with the total power flow. Linearization allows the use of superposition.

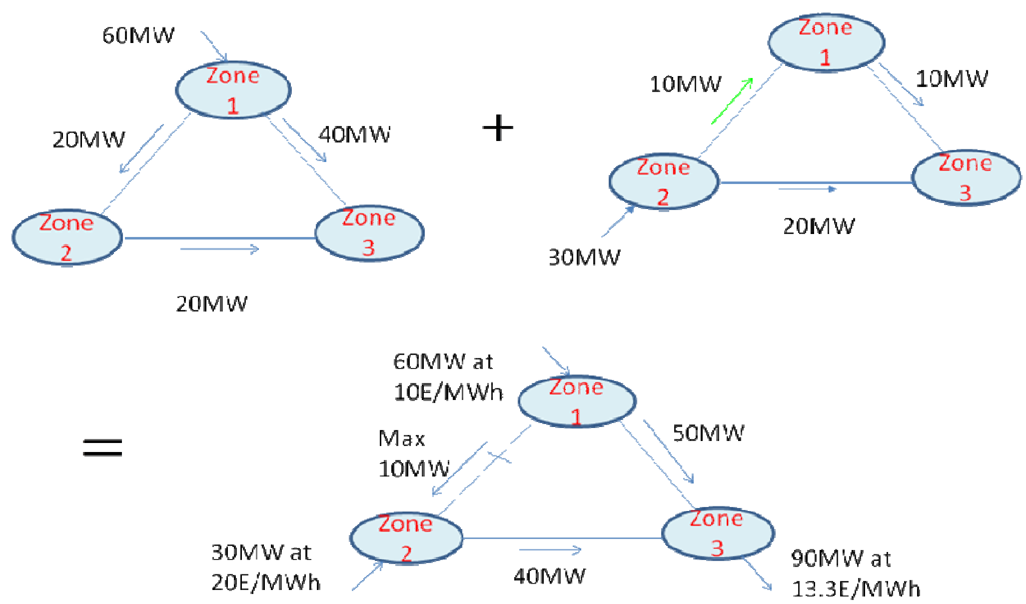

Figure 2.8 Inter-zonal power flows in the three zone network without transmission capacity limit

2.6 Conclusion

The contribution of this chapter is threefold. First, an intensive literature survey was conducted and different PTDF concepts in the literature and their corresponding network representations are categorized and presented. For each form of PTDF and network representation, an example is given with figures, tables and interpretations to show how the basic principles work. Influencing factors of nodal and zonal PTDF are pointed out after explaining the concept. Secondly, a derivation of the nodal PTDF matrix is developed as well as the relationship between nodal PTDF and transfer nodal PTDF matrices. Lastly advantages and disadvantages of aggregating the network into zonal network model and adopting zonal PTDF are discussed in comparison with nodal network model and nodal PTDF.

Chapter 3 Calculation Methods for Zonal PTDF

3.1 Introduction

In this chapter, two zonal PTDF calculation methods are discussed in detail. Before we select the typical PTDF matrix values to form a PTDF matrix family that can represent the whole year, it is important to look at the influence factors first and at the same time investigate the PTDF calculation methods and how they can be implemented.

This chapter is organized as follows. First, the New England System is briefly introduced, which the simulation throughout the master thesis is run. Then the influence factors of zonal PTDF matrices are discussed, which gives information on how to choose typical scenarios to form a PTDF matrix family in chapter 4. Later, the "classic" PTDF calculation method and GSK PTDF calculation method are investigated.

3.2 New England Model Description

The simulations for this master thesis are run on the New England test system, which is devised from the IEEE 39 bus test system.

There are 41 buses in the system, which are organized into 3 areas. The system has 12 generators, 19 loads, 34 transmission lines and 14 transformers. The topology of the network is depicted in Figure 3.1.

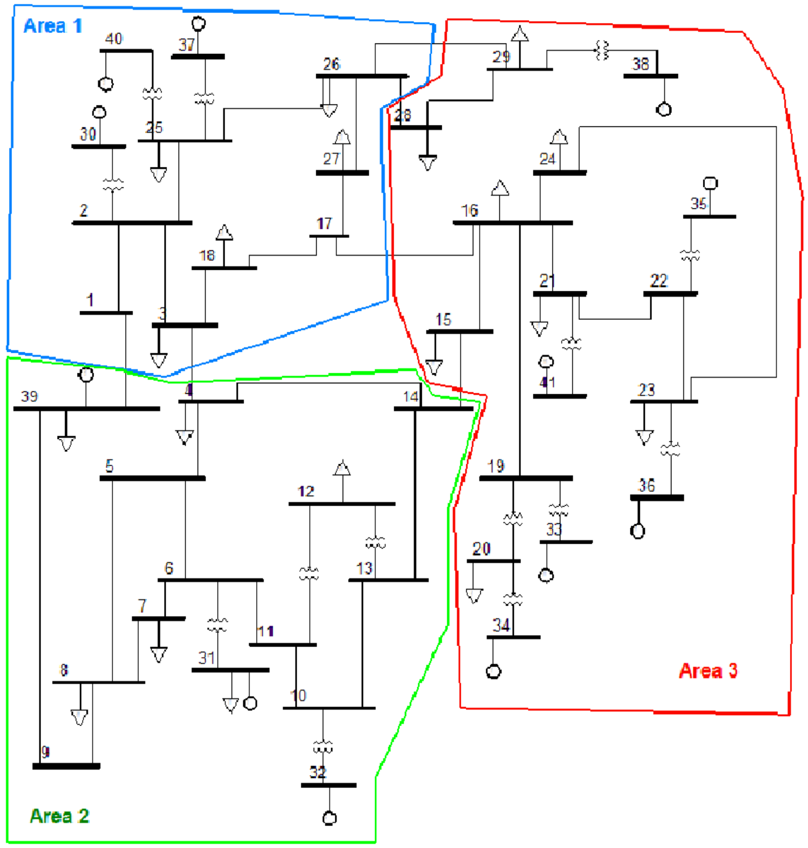

Figure 3.1 The 41-bus New England System

The three areas are connected as it follows: Area 1- Area 2 by branches 1 – 39 and 3 – 4, Area 1 – Area 3 by branches 26 – 29, 26 – 28 and 16 – 17, Area 2 – Area 3 by branch 14 – 15.

The power plant types are shown in Table 3.1. It is worth noticing that the there is one wind power plant in each area.

Power plant number	Bus number	Power plant type
1	Bus 30	Gas Turbine
2	Bus 31	Wind
3	Bus 32	Gas Engine
4	Bus 33	Gas Engine
5	Bus 34	CCGT
6	Bus 35	CCGT
7	Bus 36	CCGT
8	Bus 37	Lignite
9	Bus 38	Coal
10	Bus 39	Coal
11	Bus 40	Wind
12	Bus 41	Wind

Table 3.1 Generation type in New England system

3.3 Influencing Factors of Zonal PTDF

Duthaler [15] pointed out that zonal PTDF varies with topology, operating points and zone building. In the New England system, areas are predefined and contingency analysis related with topology change is not of interest in this master thesis, so in the following part seasonal variation and day/night variation of zonal PTDF in New England system is investigated. Area 2 is chosen to be the hub. The experience from Ercot for zone dividing is briefly introduced.

3.3.1 Seasonal Variation

In the New England system, typical winter and summer PTDF matrices are calculated using the generation and load profile of 17th January 10AM and 17th July 10AM of the study year respectively. The deviation is calculated by the equation (3.1) and the results are shown in Table 3.2:

$$Dev = \frac{PTDF_{Summer} - PTDF_{Winter}}{avg(PTDF_{Summer}, PTDF_{Winter})} \tag{3.1}$$

	AREA3>AREA2	AREA2>AREA2	AREA1>AREA2
AREA1-AREA3	4.20%	0	-22.65%
AREA1-AREA2	4.20%	0	4.93%
AREA2-AREA3	-7.94%	0	-22.65%

Table 3.2 Seasonal variation of zonal PTDF matrix

The seasonal change of generation profile and load pattern makes a substantial difference of PTDF values on certain inter-area links. Figure 3.2 depicts the load per area in winter and summer and Figure 3.3 shows the typical winter and summer generation profile per plant. In Figure 3.3, plant number 2, 11 and 12 are wind power plant and number 7 is the CCGT (combined cycle gas turbine). It is clear from the figure that in winter wind production is much higher than that of summer and so is the generation from CCGT.

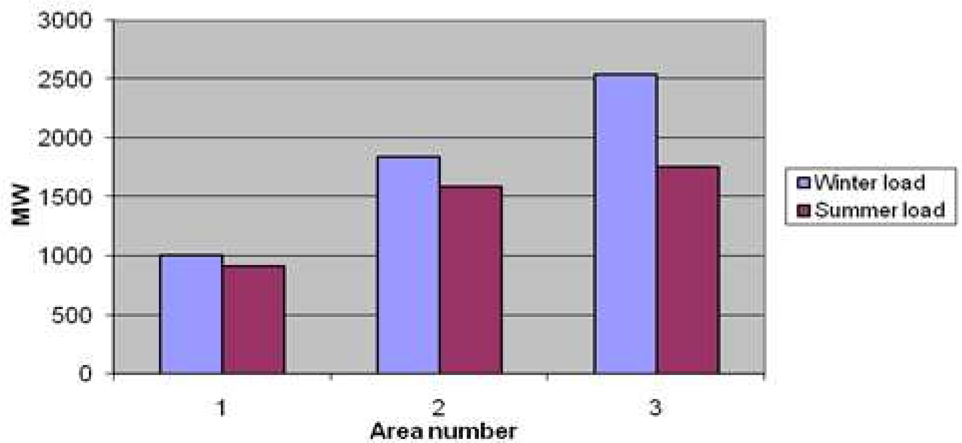

Figure 3.2 Typical winter and summer load per area

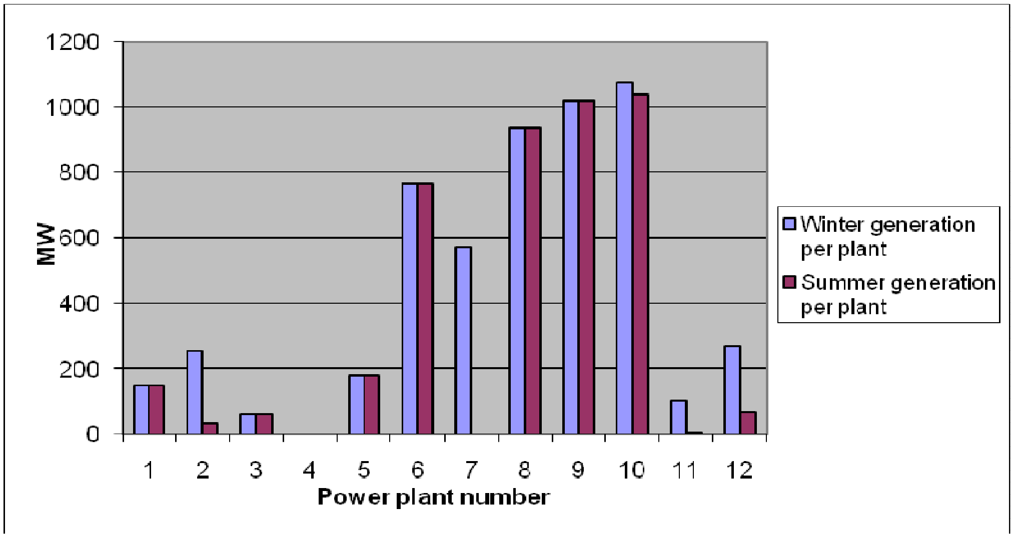

Figure 3.3 Typical winter and summer generation profile per plant

3.3.2 Day/Night Variation

In the New England system, typical winter and summer PTDF values are calculated using the generation and load profiles of 17th January 10AM and 17th January 3AM of the study year respectively. The deviation is calculated by the following equation (eq. 3.2) as shown in Table 3.3:

$$Dev = \frac{PTDF_{Day} - PTDF_{Night}}{avg(PTDF_{Day}, PTDF_{Night})} \tag{3.2}$$

	AREA3>AREA2	AREA2>AREA2	AREA1>AREA2
AREA1-AREA3	0.42%	0	-25.34%
AREA1-AREA2	0.42%	0	7.48%
AREA2-AREA3	-0.75%	0	-25.35%

Table 3.3 Day/Night variation of zonal PTDF matrix

Table 3.3 shows that the zonal PTDFs on some links are sensitive to the day and night variation. Figure 3.4 depicts the load per area in daytime and night time and Figure 3.5 shows the typical daytime and night generation profile per plant. In Figure 3.5, plant number 1 is the gas turbine and plant number 5 and 7 are CCGT. It can be seen that the load is lower at night compared with day time and the power production from gas turbine and CCGT is reduced.

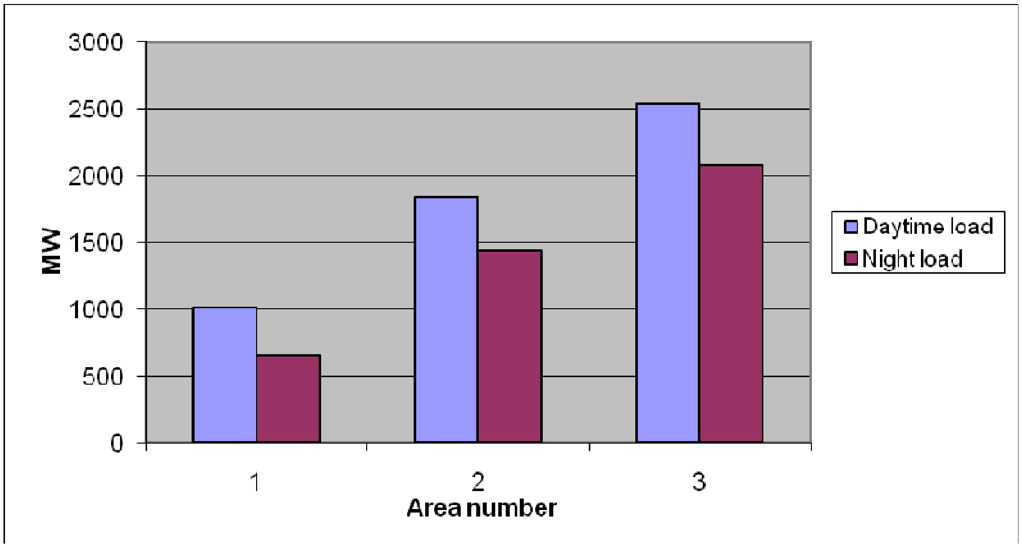

Figure 3.4 Typical daytime and night load per area

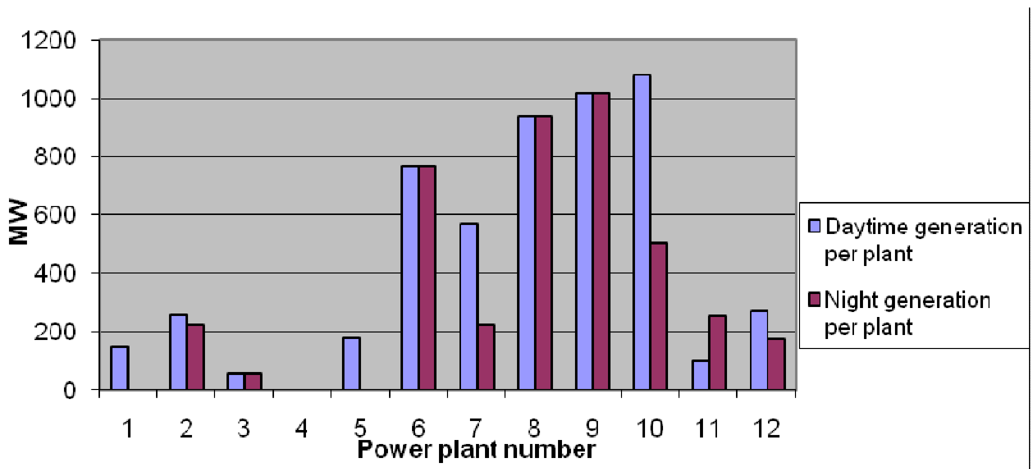

Figure 3.5 Typical winter and summer generation profile per plant

Experience in Zone Dividing from Literature (Ercot)

Ercot clusters the buses with similar nodal PTDF into a "zone". Then the nodal PTDF values across all generation in a zone are averaged according to generation weight to calculate the zonal average PTDF [16].

3.4 Classic PTDF Calculating Method

Similar to the calculation of nodal PTDF, there is a hub area assumed in zonal PTDF computation. In the classic PTDF method, the zonal network representation is used. First a reference case is chosen and load flow is performed on the reference case. Then the generation is increased by 100MW in each area except the hub area according to pre-defined generation shift method, and simultaneously the generation in the hub area is decreased 100MW each time. Power transaction of 100MW is made this way from each area to the hub area. The ratio between the changes in power transfer on each inter area link and the total power transfer (100MW) gives the zonal PTDF values:

$$PTDF^{zonal} = \frac{P_f^{'} - P_f^{ref}}{100} \tag{3.3}$$

$P_f^{'}$ Power flow on inter area link after transfer

P_f^{ref} Power flow on inter area link in reference case

New England test system is used as an example to illustrate the calculation steps. The first hour from the study year is chosen to be the reference case.

Power flow values in reference case are given by Figure 3.6:

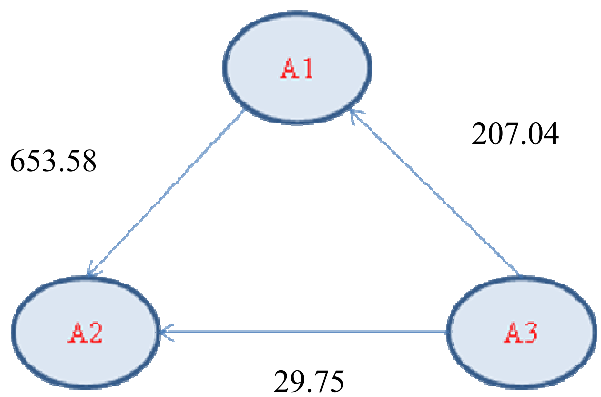

Figure 3.6 Power flow values in the reference case

First, load flow is performed after making an additional 100MW power transfer between the zone pair A1 and A2. Power flow is shown in Figure 3.7.

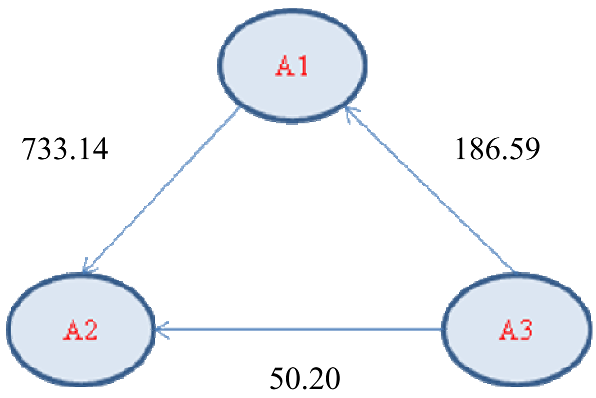

Figure 3.7 Power flow (Pf’) after 100MW transaction from A1 to A2

This new power flow is used to subtract it from the reference power flow and the ratio between differences of power flow and the requested power transaction (100MW) give the first column of Table 3.4.

The same procedure is repeated for making a power transfer from A3 to A2. The new power flow is shown in Figure 3.8.

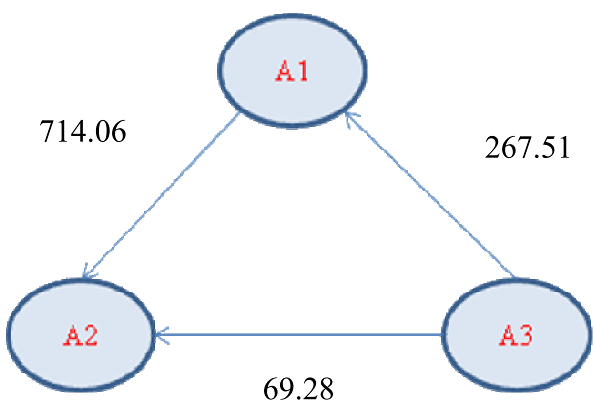

Figure 3.8 Power flow (Pf’) after 100MW transaction from A3 to A2

Zonal PTDF	Transfer A1>A2	Transfer A2>A2	Transfer A3>A2
Link A1-A2	0.7955	0	0.6047
Link A1-A3	0.2045	0	-0.6047
Link A2-A3	-0.2045	0	-0.3953

Table 3.4 Zonal PTDF calculated by classic method

3.4.1 Generation Shift Method

Vukasovic pointed out three generation shift methods in his paper [17]:

1. Increase in each generation node proportional to its production level in the reference case
2. Increase in each generation node according to merit order
3. Increase in each generation node proportional to remaining available capacity

To consider the level of information that will be available in large systems, the first generation shift method is implemented in this master thesis so that the methods and analytical conclusion presented in later chapters can be further applied to large systems.

3.4.2 Implementation

The classic PTDF calculation method is currently used by Elia, the Belgian TSO as a last step after the market calculations. The flow chart in Figure 3.9 demonstrates how classic PTDF is utilized.

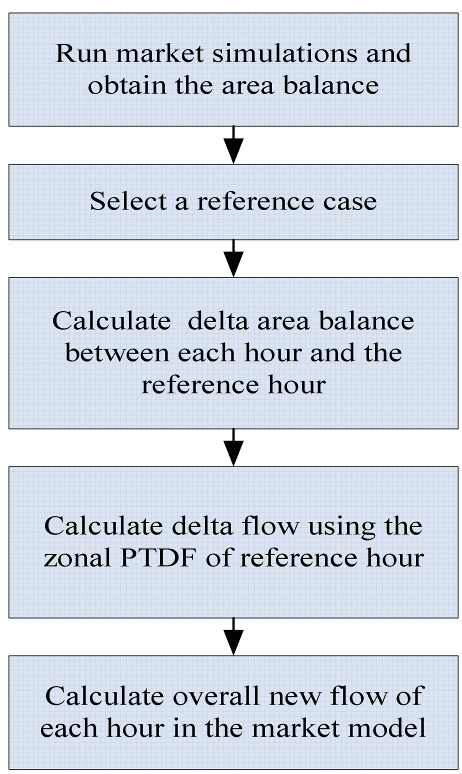

Figure 3.9 Classic PTDF implementation (post market simulations)

The New England system will be used to illustrate the steps by calculating the estimated power flow for hour 2.

Select Reference Case

First hour of the study year is chosen as a reference case here. The zonal PTDF and area balance of reference hour are presented in Table 3.4 and Table 3.5. The negative area balance value refers the area is a net importer. The positive sign indicates the area is a net exporter.

	Area 1	Area 2	Area 3
Area Balance of hour 1 (reference hour)	446	-683.25	236.70

Table 3.5 Area balance of reference case

Calculate Delta Area Balances

The delta area balance between hour 2 and reference hour is shown in Table 3.6.

	Area 1	Area 2	Area 3
Area Balance of Hour 2	425	-512	86.9
Delta Area Balance $\Delta\vec{P}_B$	21	-171.25	149.8

Table 3.6 Delta area balance between current hour (hour 2) and reference hour

Calculate Delta Flow

The Delta Flow between hour 2 and the reference hour is calculated by equation 3.4 using the matrix from Table 3.4 and result is reported in Table 3.7.

$$\Delta\vec{P}_f = PTDF^{zonal} * \Delta\vec{P}_B \tag{3.4}$$

	Link A1-A2	Link A1-A3	Link A2-A3
Delta Flow $\Delta\vec{P}_f$	107.2896	-86.5536	-63.5104

Table 3.7 Delta flow between current hour and reference hour

Calculate the Overall New Flow due to the Market Model

The estimated inter area flow at hour 2 is the sum of the delta flow which is yielded by zonal PTDF matrix and the flow in reference case. Table 3.8 reports the estimated overall power flow on inter area links at hour 2.

	Link A1-A2	Link A1-A3	Link A2-A3
Base flow in Reference Case	653.58	-207.04	-29.75
Delta Flow $\Delta\vec{P}_f$	107.29	-86.55	-63.51
Overall New Flow	760.87	-293.33	-93.26

Table 3.8 Estimated overall power flow of hour 2

The common practice of implementing zonal PTDF is to choose one operating point as the reference case for the whole year, which represents the average situation of the year. Generation and load is increased or decreased around the average operating point.

With only one reference case, the influencing factors such as seasonal variation, day/night variation are not taken into account, which inevitably brings error.

Figure 3.10 depicts the estimated power flow of the first week using hour 1 PTDF. The estimation power flow follows the reference flow closely most of the time. However, in intervals such as hour 37 to hour 48, the estimation error is larger than 60 MW, which accounts for more than 15% of the real flow. These intervals with a large estimation deviation should be identified and use a different reference case.

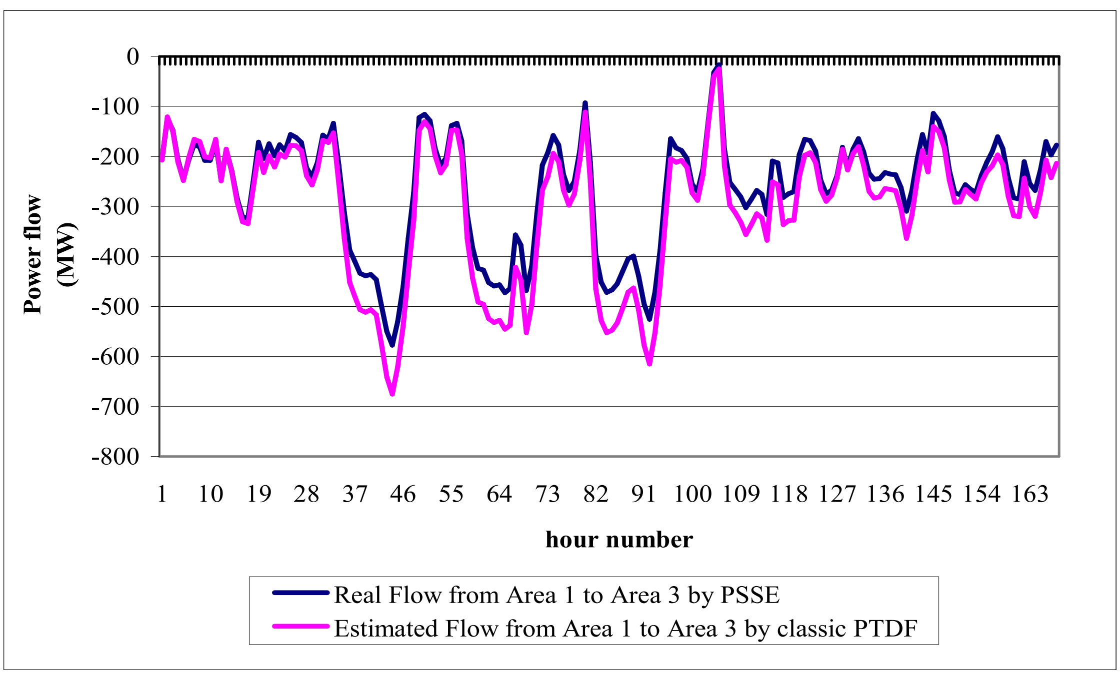

Figure 3.10 Estimation of power flow of the first week using hour 1 PTDF

3.5 GSK PTDF Calculation Method

Generation shift keys (GSK) represent the share of variation of area balance on the nodes in the zone. GSK state the forecasted merit-order in a zone and link the zonal PTDF with nodal PTDF, thus it is of interest to TSO as it bridges the difference between market model and load-flow model.

For a system with Z zones and N nodes, the dimension of the GSK is N*Z. The sum of GSK in one zone should be 1.

In this master thesis, the GSK is interpreted in net generation, which is calculated by netting generation by load, following the approach by Chao and Peck [18]. The GSK in Area 1 of the New England system is calculated in Table 3.9. When there is a 1 MW net export from Area 1, 0.42 MW is withdrawn at node 3 and 2.1 MW is injected at node 37.

Node in Area 1	**Generation (MW)**	**Load (MW)**	**Net Generation (MW)**	**Area Net Generation (MW)**	**GSK of Area1 (per unit)**
1	0	0	0	447	0.00
2	0	0	0	447	0.00
3	0	186.1	-186.1	447	-0.42
17	0	0	0	447	0.00
18	0	91.3	-91.3	447	-0.20
25	0	129.4	-129.4	447	-0.29
26	0	80.3	-80.3	447	-0.18
27	0	162.4	-162.4	447	-0.36
30	0	0	0	447	0.00
37	936.04999	0	936.04999	447	2.10
40	160	0	160	447	0.36

Table 3.9 GSK of the nodes in Area 1, hour 1

3.5.1 Relationship between Zone-to-line PTDF and Nodal PTDF

For the computation of GSK PTDF, the zone-to-line network representation is used and zonal PTDF is obtained by aggregating the zone-to-line PTDF values.

Equation 3.5 links the nodal PTDF and zone-to-line PTDF.

$$\begin{bmatrix} ptdf_{inter-line1}^{zone\ A} \cdots ptdf_{inter-line1}^{zone\ Z} \\ \vdots \quad \cdots \vdots \\ ptdf_{inter-line\ p}^{zone\ A} \cdots ptdf_{inter-line\ p}^{zone\ Z} \end{bmatrix} = \\ = \begin{bmatrix} ptdf_{inter-line1}^{node\ 1} \cdots ptdf_{inter-line1}^{node\ n} \\ \vdots \quad \cdots \vdots \\ ptdf_{inter-line\ p}^{node\ 1} \cdots ptdf_{inter-line\ p}^{node\ n} \end{bmatrix} \times \begin{bmatrix} gsk_{node1}^{zoneA} \cdots gsk_{node\ 1}^{zone\ Z} \\ \vdots \quad \cdots \vdots \\ gsk_{node\ n}^{zone\ A} \cdots gsk_{node\ n}^{zone\ Z} \end{bmatrix} \tag{3.5}$$

The steps of calculating GSK PTDF method will be explained using New England test system as an example. In the New England system, the line from node 1 to node 39 and the line from node 3 to node 4 are the inter area lines from area 1 to area 2. The nodal PTDF of nodes in area 1 on the inter area lines and GSK of nodes in Area 1 are shown in Table 3.10.

We can compute the zone-to-line PTDF with equation 3.6 and equation 3.7:

$$PTDF_{Area1,line1-39}^{zone-to-line} = PTDF_{nodes\ in\ Area1,line1-39}^{nodal} * GSK_{Area1} = 0.6673 \tag{3.6}$$

$$PTDF_{Area1,line3-4}^{zone-to-line} = PTDF_{nodes\ in\ Area1,line3-4}^{nodal} * GSK_{Area1} = 0.2519 \tag{3.7}$$

Then the zonal PTDF of Area 1 on the equivalent link between Area 1 and Area 2 is calculated by equation 3.8:

$$PTDF_{Area1,linkA1-A2}^{zonal} = PTDF_{Area1,line1-39}^{zone-to-line} + PTDF_{Area1,line3-4}^{zone-to-line} = 0.9192 \tag{3.8}$$

Node number in Area1	Nodal PTDF(1-39) of Area1	Nodal PTDF(3-4) of Area1	GSK of Area1
1	0.8523	0.1015	0.00
2	0.6095	0.2685	0.00
3	0.5366	0.3529	-0.4168
17	0.5225	0.167	0.00
18	0.5279	0.2379	-0.2045
25	0.5993	0.2565	-0.2898
26	0.5607	0.2115	-0.1798
27	0.5432	0.1911	-0.3637
30	0.6095	0.2685	0.00
37	0.5993	0.2565	2.0961
40	0.5993	0.2565	0.3583

Table 3.10 Nodal PTDF of area 1 on the inter area lines from area 1 to area 2 and GSK of area 1

Similarly, the zonal PTDF matrix can be calculated using the method described above and Table 3.11 reports the GSK zonal PTDF of hour 1.

Zonal PTDF	Area 1	Area 2	Area 3
LinkA1-A2	0.9192	0.0631	1.2116
LinkA1-A3	0.0805	-0.0631	-1.2116
LinkA2-A3	-0.0805	0.0631	0.2149

Table 3.11 GSK PTDF of hour1

The estimated power flow is obtained by equation 3.9:

$$P_f = PTDF^{zonal} * P_B \tag{3.9}$$

P_B is the area net generation .

The estimated power flow on each inter area link at hour 1 is compared with the real power flow calculated in PSSE in Table 3.12.

	Link A1-A2	Link A1-A3	Link A2-A3
Estimated power flow $P_f{}'$	654.1205	-207.653	-28.2467
Real power flow P_f **by PSSE**	653.5309	-206.98	-29.7182

Table 3.12 Estimation of hour1 power flow using the GSK PTDF

In Figure 3.11, we use the GSK PTDF of the first hour in the first week of the study year and apply equation 3.9 to obtain the power flow on inter area links.

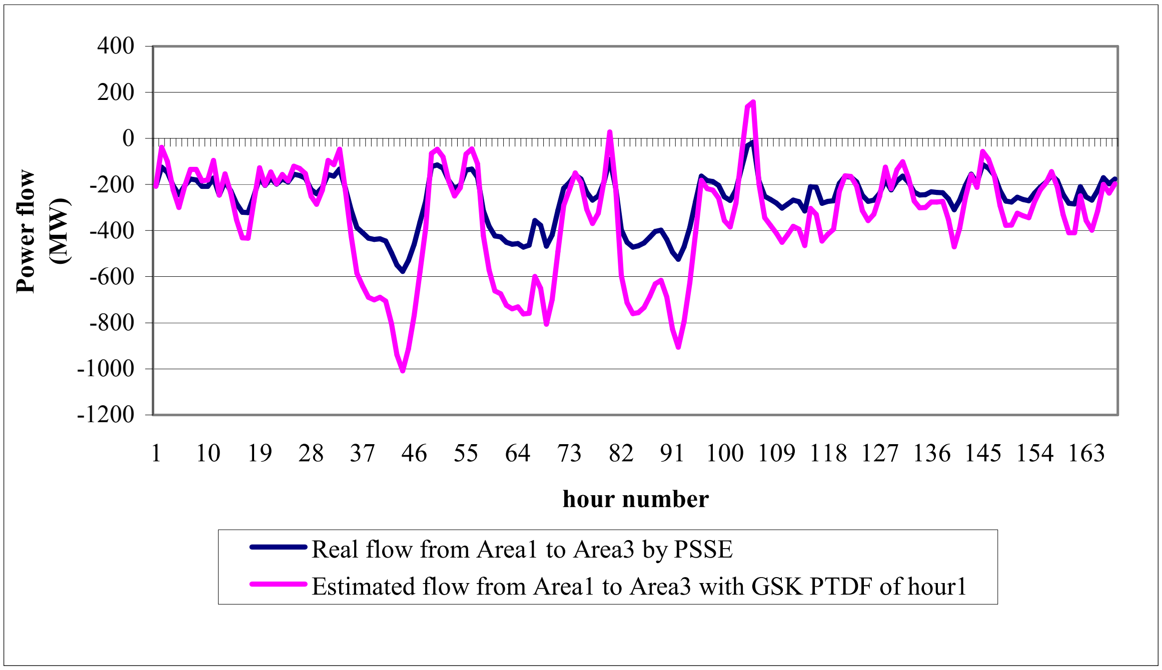

Figure 3.11 Weekly estimation of power flow using GSK PTDF of hour 1

3.6 Conclusion

The contribution of this chapter is twofold. First, when we investigate the PTDF matrix variation in the New England system brought by season change, day/night variation. Then we get further insight into how the seasonal changes, the day/night variation influence the power production by plant type, and hence also they influence the PTDF factors. Then in the PTDF calculation method, not only are the calculation steps illustrated with New England system, but also the way of implementing the two types of PTDF matrices is presented in detail.

Figure 3.10 and Figure 3.11 show the estimation power flow of week 1 in the study year by using classic PTDF and GSK PTDF of the first hour. It is clear from these two figures that GSK PTDF method gives a larger error in this case. As Ehrenmann and Smeers [7] pointed out, that the GSK method implicitly implies a nodal view of the network, which requires the UC-ED result in a nodal level to calculate the GSK PTDF matrix. It will bring substantial computation difficulty. Moreover, this information is not always available in a larger system. Hence, for implementation in the large systems, the classic PTDF method is chosen to be used further this master thesis for creating the dynamic PTDF matrix family.

Chapter 4 Data Analysis Methodology and Dynamic PTDF Family Identification

4.1 Introduction

The main driver of this research is a more accurate representation of the transmission grid within the market simulations. Only using NTC values without considering the inter dependencies between different links can cause congestion in the real grid. That is why the PTDF representation of the cross border links is a step forward. Moreover, when one looks at the time frame of 1 year, it is also incorrect to consider only one PTDF matrix for all the hours of the year, and more representative PTDF matrices have to be used for different hours.

In chapter 3, different network representation and their corresponding forms of PTDF matrix are introduced. In this chapter, dynamic PTDF family is formed such that several PTDF matrices of typical scenarios are assigned to represent all hours in the time interval. The methodology to select typical scenarios to form dynamic PTDF family is developed and the effect of implementing the PTDF matrices of typical scenarios is assessed and compared with the case of using only one PTDF matrix.

The chapter is organized as follows. Firstly, the mathematical methods that are implemented to form dynamic PTDF families are introduced. Then the typical scenarios selection process is presented and the rule of choosing one of the typical scenarios to represent a random hour is defined. In this chapter, New England system is used to demonstrate the methodology.

An overview of the procedures to form dynamic PTDF family is shown in Figure 4.1:

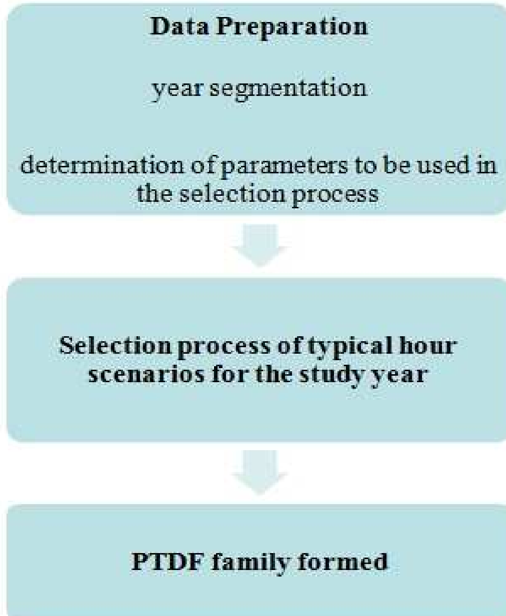

Figure 4.1 Overall procedures of forming the dynamic PTDF family

4.2 Input Data Preparation for Data Analysis

4.2.1 Hour Number Segmentation

Considering the influence factors discussed in chapter3 and the size of data set that is required to apply the partition method, we first segment the hours of the year and find typical PTDF values for each interval.

In section 4.3 and section 4.4, New England test system as shown in Figure 3.1 is used as study case to demonstrate the methodology. The data set of the balanced hour points-hourly generation equals load in the initial UC-ED outcome-of the study year, is first divided as weekday and weekend groups. Then different time of the day is defined according to the experience at Elia the Belgian TSO and later four seasons are used to segment the data. In the end, we obtain 24 intervals.

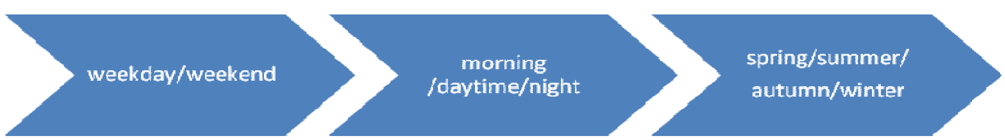

Figure 4.9 Segmentation of the hours in the whole year

Experience in this master thesis shows a significant improvement of accuracy of the estimated power flow with typical PTDF values chosen by applying k-medoids method to data from the intervals rather than the whole year. In the following part, we take the interval of weekday/daytime/autumn to demonstrate the data analysis method.

4.2.2 Selection of Parameters for Data Analysis Using Linear Regression

Linear regression is performed on the study case. With the first hour of the year chosen as the reference case, the inter-area power flow for the rest of hours in the year is estimated. The difference between the estimated power flow in each interval defined in 4.6.1 and the first hour power flow is used as the dependent variable y.

The area load, area wind and net area generation produced by Powrsym4 without implementing PTDF are used as the independent variable X .

The correlation coefficient between the incremental power flow and the area load, area wind and net exchange without implementing PTDF in the interval of weekday/daytime/winter is 0.9989. Correlation coefficients in other intervals are at the same level between 0.99 and 1. This correlation coefficient demonstrates a strong correlation between chosen y and X . That is, area load, are wind and net area generation without implementing PTDF provide very useful information to estimate incremental power flow compared with reference case. Thus, area load, area wind and net area generation are chosen to form the data set for k-medoids method to select the hours with typical PTDF values.

Linear regression coefficients give the insight of how much each of the independent variables influence the dependent variables and it is used to define the distance between independent variables of typical PTDF hours and the other hours with the purpose of locating the outliers.

Table 4.1 shows the regression coefficient β of the incremental power flow and area load, area wind and net exchange in the market model (without implementing PTDF logic) in the New England case for the interval weekday/daytime/autumn.

Table 4.1 reveals that the Net area balance is the dominant influence factor to estimate difference between estimated power flow and the reference flow. W refers to wind, LA load area and NAB net area balance.

Independent variable x	**W1**	**W2**	**W3**	**LA1**	**LA2**	**LA3**	**NAB1**	**NAB2**	**NAB3**
regression parameter β	0.0027	0.0009	0.0066	0.0265	0.0069	0.0118	2.8474	2.5956	1.9866

Table 4.1 Regression parameters of the interval weekday/daytime/autumn

4.3 Typical Hours Selection Process and Study Case

In each interval, three selection stages are implemented. The input and output of data process are shown in Figure 4.10. UC-ED should first be performed in the market model to obtain the net area balance. It should be noticed that the complete three steps are demonstrated for New England test system in the time interval where outlier level of the input data is high. For the input data set where the data objects are close to each other, it is not necessary to perform the complete three stages. Data process can exit after stage 2 if all distances obtained in this stage are below the threshold predefined.

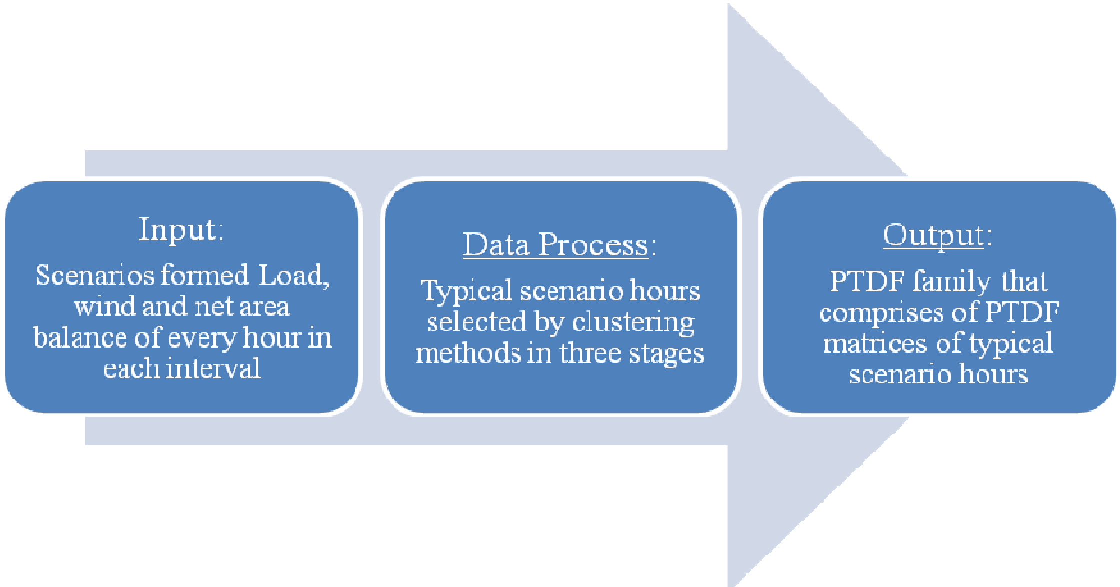

Figure 4.10 Input and output of the data selection process

To get the typical PTDF values, three selection stages are implemented in each interval. (See section 4.6) New England test system is used to demonstrate the three steps.

Stage 1: Steps in Stage 1 is illustrated in Figure 4.11. In the first four steps from Figure 4.11, the k-medoids partition method is implemented in combination with mean silhouette value and the latter is used as an aid to determine optimum number of clusters. The medoids returned by k-medoids method are typical area load, area wind and net generations and the indexes of the medoids are the typical hour numbers.

In step 5, for each hour i in the interval, the distance between hour i and typical PTDF hour j is calculated by using equation 4.7.

$$d(i,j) = \sqrt{(\beta_1 x_{i1} - \beta_1 x_{j1})^2 + (\beta_2 x_{i2} - \beta_2 x_{j2})^2 + \cdots (\beta_n x_{i9} - \beta_n x_{j9})^2} \qquad (4.7)$$

The area load, area wind and area net exchange obtained from the Powrsym4 UC-ED simulations without implementing PTDF form the independent variables $x_{i1} \cdots x_{i9}$. The coefficients $\beta_1 \cdots \beta_9$ are obtained from the linear regression analysis in Table 4.1 corresponding to independent variables $x_1 \cdots x_9$. Regression coefficients $\beta_1 \cdots \beta_9$ weigh the independent variables in distance measures. In the interval of weekday/daytime/autumn, five typical PTDF hours are chosen in stage 1 by k-medoids method in combination with silhouette value. For each hour i, the distances with respect to the five typical PTDF hours are calculated separately and the typical PTDF hour with the smallest distance is chosen to represent hour i, in step 6.

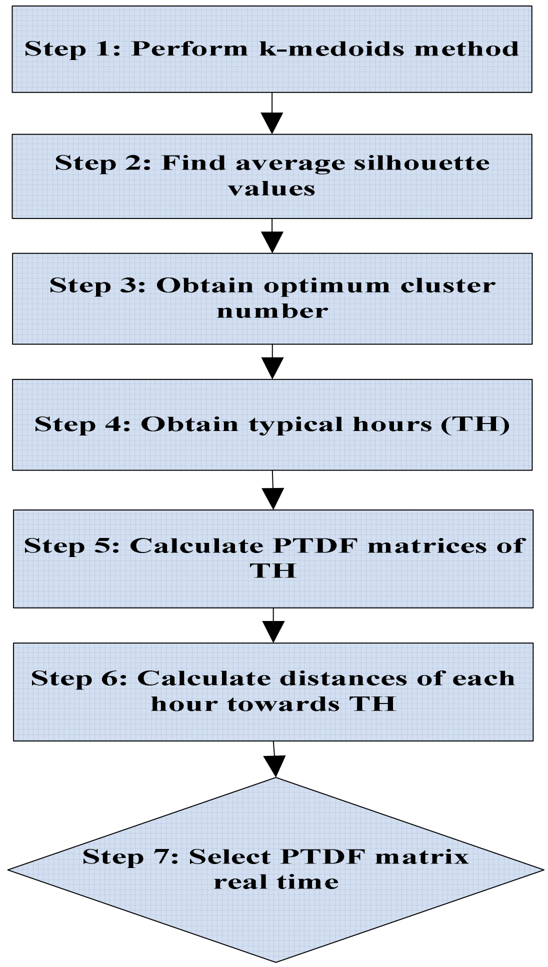

Figure 4.11 Steps in Stage 1

In the interval of weekday/daytime/autumn, the silhouette value reaches maximum when the number of clusters is 5. The k-medoids algorithm returns the five typical scenarios shown in Table 4.2.

x_j / Scenarios	W1 (MW)	W2 (MW)	W3 (MW)	L1 (MW)	L2 (MW)	L3 (MW)	NA1 (MW)	NA2 (MW)	NA3 (MW)
TS1	253	182	341	1052	1896	2283	287	-575	288
TS2	251	235	312	1043	1920	2117	-642	-548	1190
TS3	88	89	110	946	1776	1942	699	-311	-388
TS4	182	70	168	986	1759	1895	282	-551	269
TS5	224	86	218	723	1327	1544	436	-905	469

Table 4.2 Five typical scenarios selected in stage 1

The independent variables of an arbitrary hour are shown in Table 4.3.

Scenario	WA1 (MW)	WA2 (MW)	WA3 (MW)	LA1 (MW)	LA2 (MW)	LA3 (MW)	NAB1 (MW)	NAB2 (MW)	NAB3 (MW)
Independent variable x_i	5.0	54	44	632	992	1106	309	-603	294

Table 4.3 Variables of an arbitrary hour

Distances between the hour in Table 4.3 and the five typical scenario hours are calculated using equation 4.7 and regression parameter in Table 4.1are: 96, 3244, 1907, 164, 933.

As typical scenario 1 has the smallest distance with the arbitrary hour, PTDF matrix of typical scenario 1 is used to represent this hour.

Estimation of the power flow of this hour is calculated using the PTDF matrix of typical scenario 1, area balance of typical scenarios 1 and the incremental area balance between the current hour and the typical scenario hour. (See chapter 3)

Similarly, estimated power flow using the five typical PTDF matrices are calculated for each hour in the weekday/daytime/autumn interval and compared with the real power flow obtained in PSSE. Figure 4.12 shows the real power flow of the interval weekday/daytime/autumn calculated in PSSE. For calculation of absolute errors and relative errors, the flow between area 1 and area 2 is used.

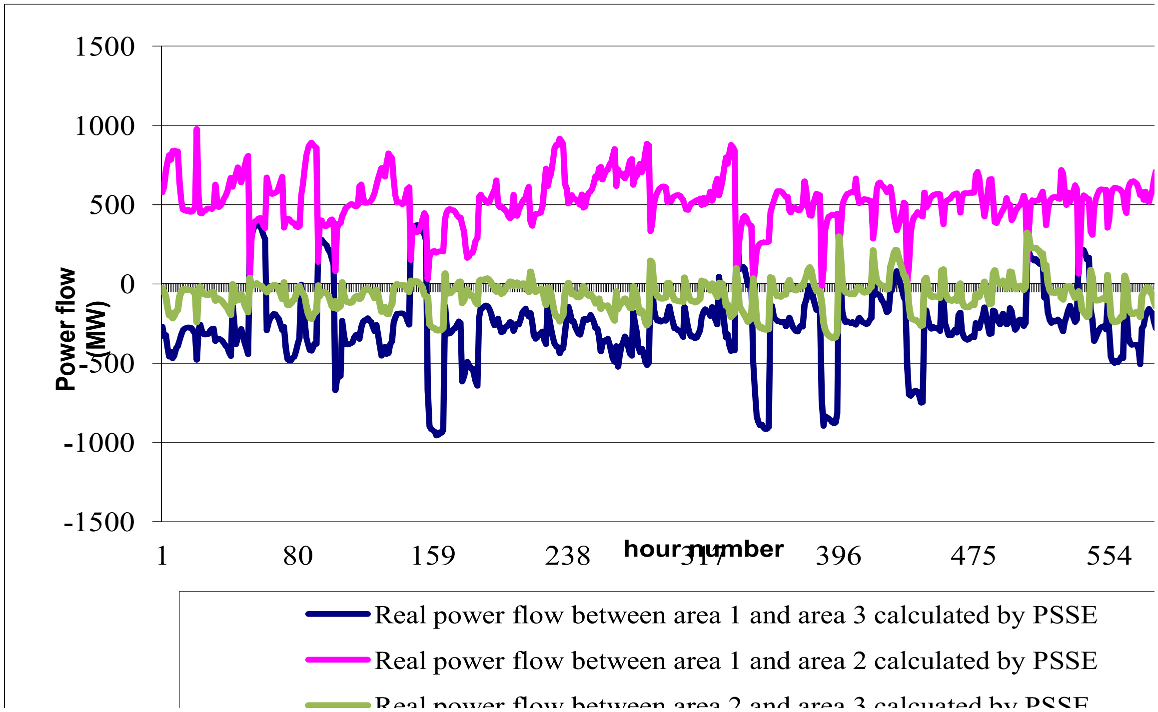

Figure 4.12 Real power flow of the interval weekday/daytime/autumn calculated in PSSE

Figure 4.13 and Figure 4.14 show the error of estimated power flow with only one fix PTDF matrix and the estimated flow with 5 representative PTDF matrices selected from stage 1.

The absolute error r is calculated using equation 4.8:

$$E_{abs} = \hat{P}_f - P_f \tag{4.8}$$

Where

$\hat{P}_f$ is the estimated power flow using PTDF matrix

P_f is the real power flow calculated in PSSE

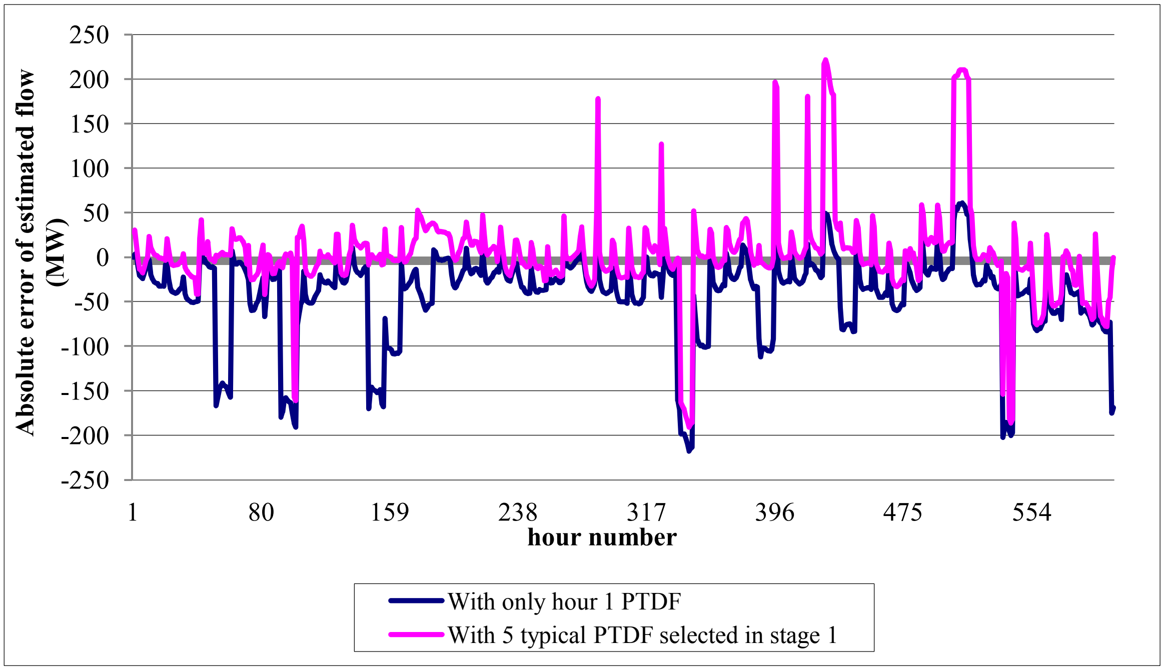

Figure 4.13 Absolute error of estimated power flow after stage 1

When we take power flow of the most congested link to evaluate the relative error, it can be expressed according to equation 4.9. Only the power flow that is larger than 200MW is considered, since the light loaded situation is not of our interest.

$$E_{rel} = \frac{\hat{P}_f - P_f}{P_f} \times 100\% \tag{4.9}$$

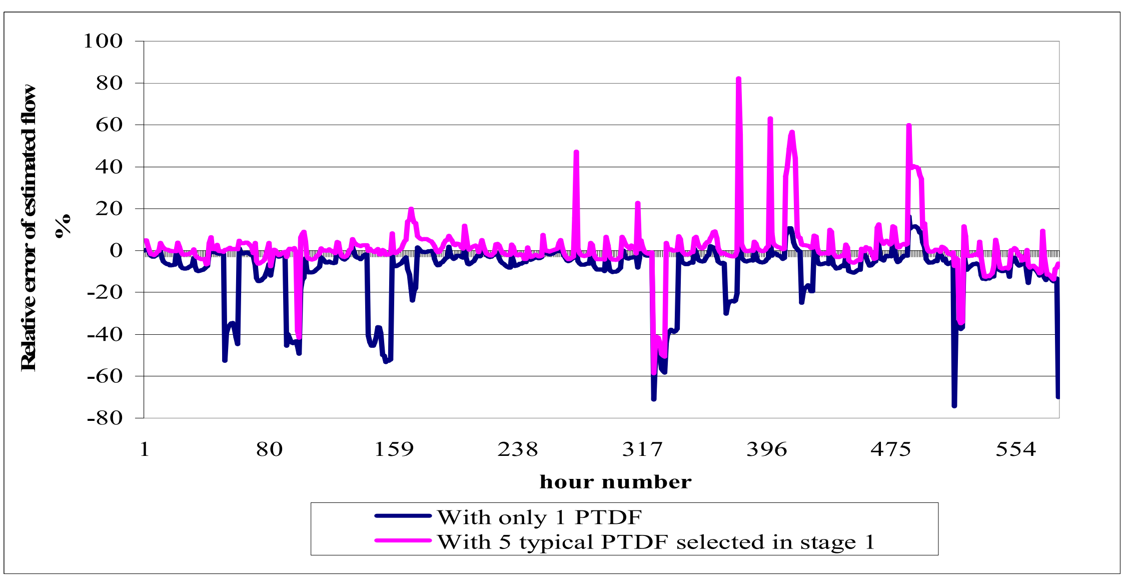

Figure 4.14 Relative error of estimated power flow after stage 1

As illustrated in Figure 4.13 and Figure 4.14, for some intervals power flow estimation error is largely reduced. On the other hand, the error rises sharply for several hours, which suggests post processing is necessary after stage 1. Nonetheless, the five typical scenarios in stage1 establish a base to measure the distance of the load, wind and net exchange values between the typical scenario hours and the current hour.

Stage 2: Subzone division

The large error points do not always coincide with large values of the minimum distance with typical scenario hours computed in stage 1. Before we deal with the outliers from the data perspective, the physical layer is analyzed to tackle this inconsistency. Buses with similar nodal PTDF in each area are grouped into subareas. The introduction of subareas gives a more precise description of the system and variables in distance calculation are adjusted accordingly.

It is obvious from Table 4.4 that generation buses in area1 have similar Nodal PTDF on the inter-area lines, which implies that decrease or increase of the same amount of generation in any of the buses in area 1 will have similar effect on power flow on the inter-area links. Thus it is logical to group these buses in one area.

Link	Line	Bus30	Bus37	Bus40
A1-A2	**1-39**	0.6095	0.5993	0.5993
A1-A2	**3-4**	0.2685	0.2565	0.2665
A1-A2	**26-29**	0	0	0
A1-A3	**26-28**	0	0	0
A1-A3	**17-16**	0.122	0.1442	0.1442
A2-A3	**14-15**	-0.122	-0.1442	-0.1442

Table 4.4 Nodal PTDF of Area 1

In Table 4.5, we can also see that all generation buses in area 2 have similar Nodal PTDF on the inter-area lines, therefore area 2 is treated as a single integrate area.

Link	Line	Bus30	Bus37	Bus40
A1-A2	**1-39**	0.3977	0.4214	0
A1-A2	**3-4**	-0.27	-0.2536	0
A1-A2	**26-29**	0	0	0
A1-A3	**26-28**	0	0	0
A1-A3	**17-16**	-0.1278	-0.1678	0
A2-A3	**14-15**	0.1278	0.1678	0

Table 4.5 Nodal PTDF of Area 2

In Table 4.6, it is interesting to notice that bus 38 has a large PTDF value on the inter area link between area 1 and area 3, while the rest of the generation buses have no influence on this link. In other words, bus 38 is the only generation bus in the area that can bring power flow between area 1 and area 3, via lines 26-29 and 26-28. Besides, the PTDF values of Bus 38 on the inter-area link between area 1 and area 3 is much smaller than the rest of the buses, which indicates that the same amount of power injection or withdrawal change in bus 38 will have less influence on the inter area link between area1 and area3 compared with other buses.

Link	Line	Bus 33	Bus 34	Bus35	Bus36	Bus38	Bus 41
A1-A2	**1-39**	0.5061	0.5061	0.5061	0.5061	0.5607	0.5061
A1-A2	**3-4**	0.0776	0.0776	0.0776	0.0776	0.2115	0.0776
A1-A2	**26-29**	0	0	0	0	-0.5	0
A1-A3	**26-28**	0	0	0	0	-0.5	0
A1-A3	**17-16**	-0.5837	-0.5837	-0.5837	-0.5837	0.2278	-0.5837
A2-A3	**14-15**	-0.4163	-0.4163	-0.4163	-0.4163	-0.2278	-0.4163

Table 4.6 PTDF of Area 3

From the discussions above, it is reasonable to divide subareas in area3 and keep area1 and area2 as they are.

The resulting network representation is shown in Figure 4.15.

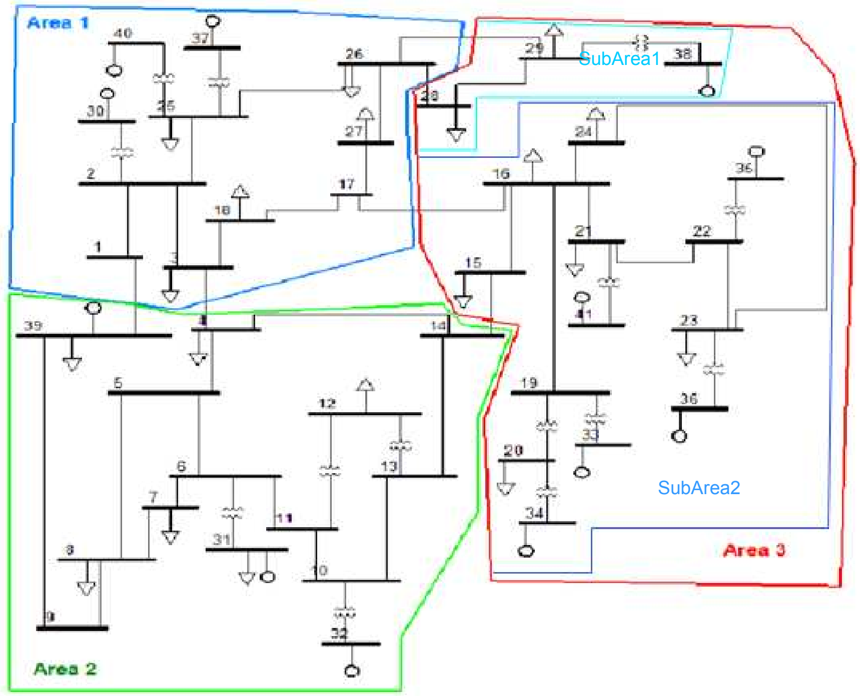

Figure 4.15 The 41-Bus New England network with subareas

New distance defined to select typical PTDF

With the newly added subareas, variable dimension in the distance calculation needs to be increased. Yet the distance calculation should not be too complicated for application in large systems. In the case of weekday/daytime/winter, we can see from Table 4.1 in Section 4.3.2 that the regression parameter of net exchange is dominant, so the subarea values are specified only in the category of net exchange. The five typical scenario hours chosen in stage 1 with subarea variables is shown in Table 4.7.

variable x_i	WA1 (MW)	WA2 (MW)	WA3 (MW)	LA1 (MW)	LA2 (MW)	LA3 (MW)	NAB1 (MW)	NAB2 (MW)	SAB1 (MW)	SAB2 (MW)
TS1	253	182	341	1052	1896	2283	287	-575	596	-308
TS2	251	235	312	1043	1920	2117	-642	-548	627	563
TS3	88	89	110	946	1776	1942	699	-311	-359	-28
TS4	182	70	168	986	1759	1895	282	-551	668	-400
TS5	224	86	218	723	1327	1544	436	-905	733	-264

Table 4.7 Five typical scenario hours chosen in stage 1 with subarea variables

The variables of the arbitrary hour in Table 4.3 are adjusted as shown in Table 4.8.

variable x_j	**WA1 (MW)**	**WA2 (MW)**	**WA3 (MW)**	**LA1 (MW)**	**LA2 (MW)**	**LA3 (MW)**	**NAB1 (MW)**	**NAB2 (MW)**	**SAB1 (MW)**	**SAB2 (MW)**
TS1	5.0	54	44	632	992	1106	309	-603	756	-462

Table 4.8 Variables of the arbitrary hour with subzone divided

Distances between the hour in Table 4.8 and the five typical scenario hours are recalculated using equation 4.7. The new distances are: 451, 3402, 2730, 265, 952. Typical scenario 4 has the smallest distance with this hour, so PTDF matrix of typical scenario 4 becomes the new representative PTDF matrix for this hour in stage 2.

Estimation power flow for each hour in the interval of weekday/daytime/autumn with the same five typical PTDF matrices but updated distances are calculated again and compared with the real power flow obtained in PSSE.

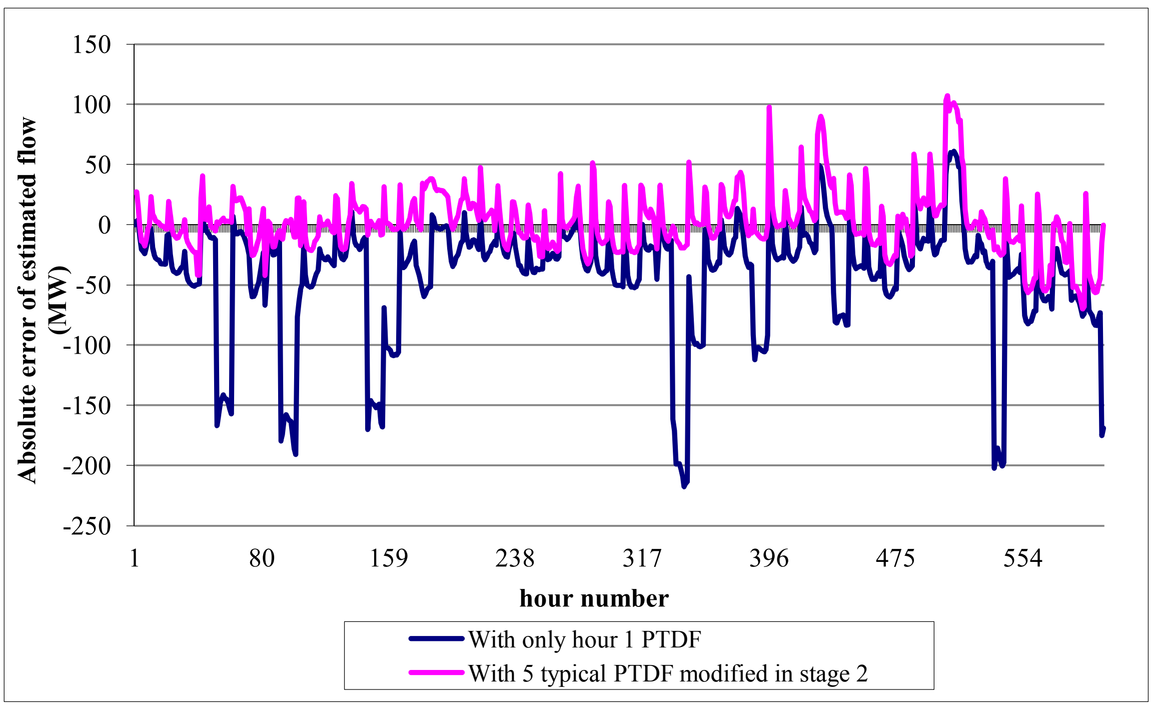

Figure 4.16 Absolute error of estimated power flow after stage 2

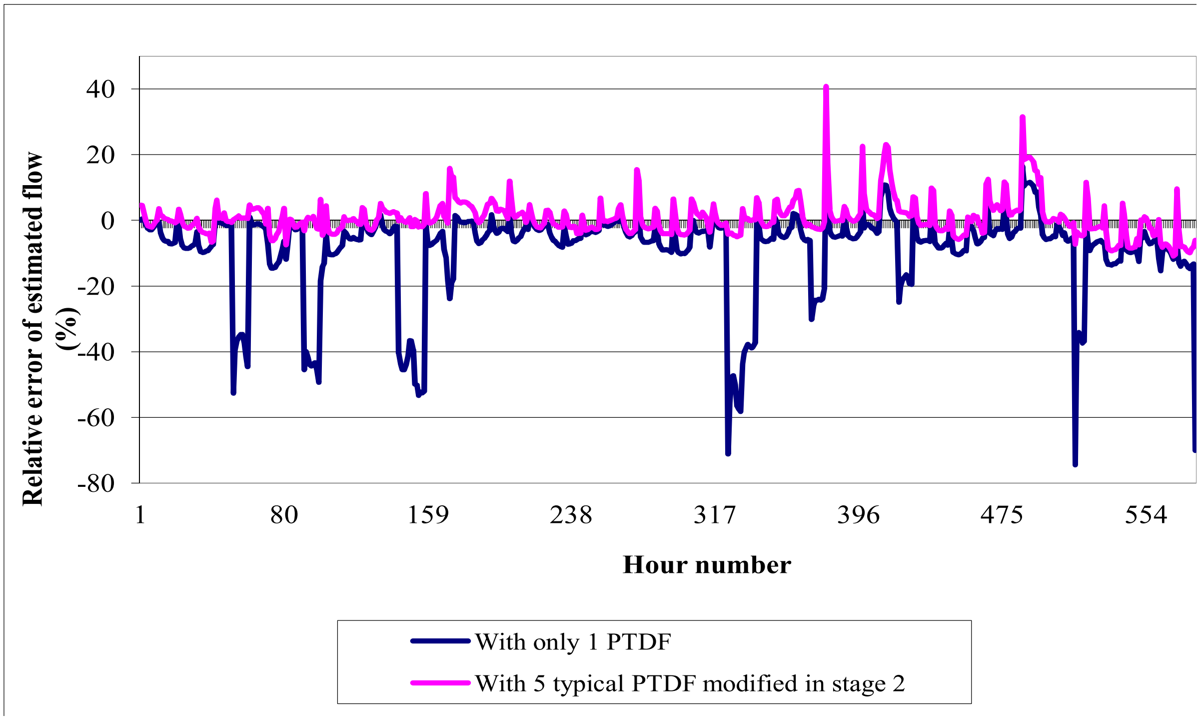

Figure 4.17 Relative error of estimated power flow after stage 2 with the real power flow value larger than 200MW

As can be seen in Figure 4.16 and Figure 4.17, subzone division in stage 2 facilitates a substantial improvement in the accuracy of estimated power flow. The peak value of absolute error is halved compared with using only one PTDF matrix.

Stage 3:

Since the accuracy of power flow estimation is very sensitive to outliers, the second mathematical selection process is aimed at finding the representative PTDF hour number among the outliers to further alleviate the spikes in Figure 4.16.

For the interval of weekday/daytime/autumn, hours with minimum distance larger than 1500MW towards the five typical scenario hours in Table 4.7 are taken as outliers to form the new data set for the k-medoids method and silhouette evaluation. The steps in Figure 4.11 are followed with the outlier data set and four typical scenario hours are selected as shown in Table 4.9.

variable x_i	WA1 (MW)	WA2 (MW)	WA3 (MW)	LA1 (MW)	LA2 (MW)	LA3 (MW)	NAB1 (MW)	NAB2 (MW)	SAB1 (MW)	SAB2 (MW)
TS6	262	125	109	748	1532	1515	-335	-269	-134	738
TS7	263	263	264	938	1342	2236	411	60	605	-1076
TS8	270	154	262	982	1891	1976	374	-598	-365	591
TS9	118	315	128	1009	1884	2362	666	-309	582	-939

Table 4.9 Four typical scenarios among the outliers

Then we merge the typical scenarios in Table 4.9 with the ones in Table 4.7. In total, there are nine typical scenarios for the interval of weekday/daytime/autumn. For each hour in the interval, nine distances with respect to the typical scenarios are calculated using equation 4.7. The PTDF matrix of the typical scenario with smallest distance is used to represent the hour.

Power flow is estimated for each hour in the interval of weekday/daytime/autumn with the nine typical scenarios and the absolute error is shown in Figure 4.18. For relative error, again we take the hour points with the real power flow value larger than 200MW and the result is illustrated in Figure 4.19.

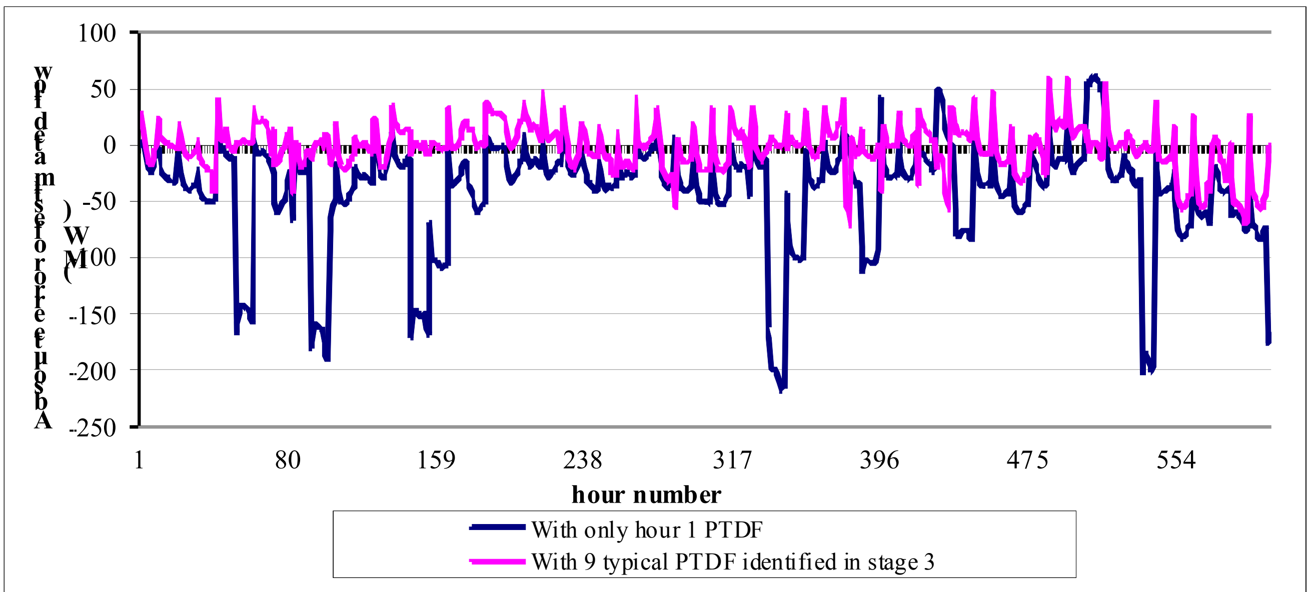

Figure 4.18 Absolute error of estimated power flow after stage 3

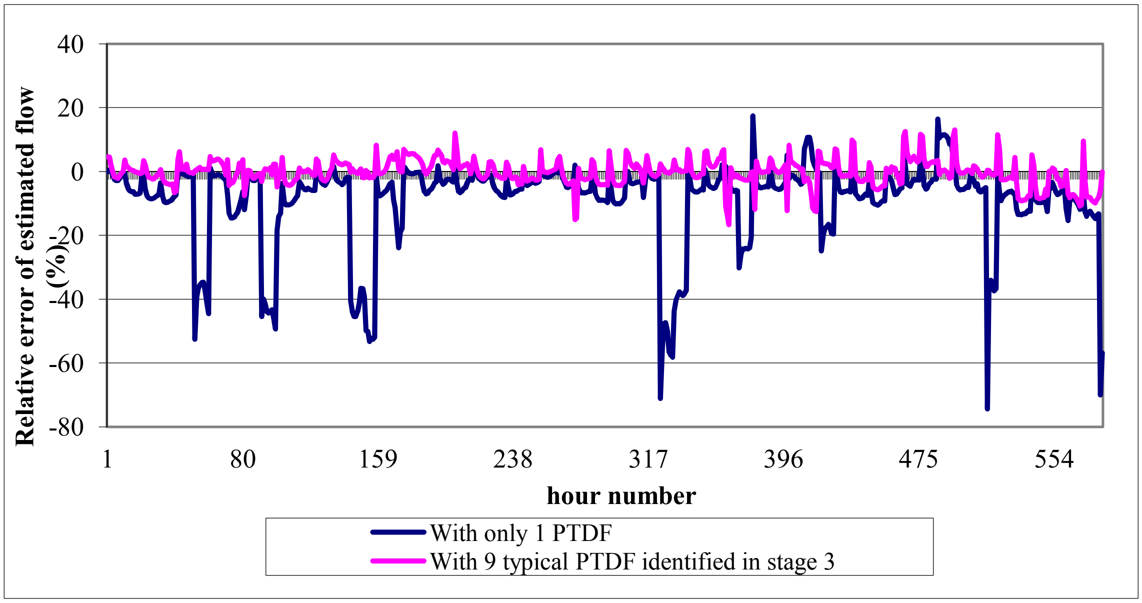

Figure 4.19 Relative error of estimated power flow after stage 3 with the real power flow value larger than 200MW

Figure 4.18 and Figure 4.19 demonstrate another leap in the accuracy of power flow estimation. The peak value of power estimation error is further reduced to around 50MW with the nine representative PTDF matrices. There is about 67% reduction of the peak value of absolute error compared with using only one PTDF matrix. The average error of the estimated flow reduced from 16.9% with one fixed PTDF matrix to 3.7% when we adopt the PTDF family.

4.4 Conclusions

Currently, the common practice for PTDF matrix application is to use only one PTDF matrix all the time for power flow estimation, which may lead to substantial estimation error of power flows in the interconnections. The large error estimation of power flow is a waste of efficiency and underestimation in heavily loaded network can hamper system security. Thus in this master thesis we develop the PTDF family concept to find a representative PTDF matrices that can be dispatched by market model for a random hour by assessing the input data.

The main contribution of this chapter is that the methodology is developed for selecting typical scenarios to form dynamic PTDF family and at the same time, the rule is defined to choose a representative PTDF matrix for a random hour from the PTDF matrix family. A dynamic PTDF family formation procedure is demonstrated with New England test system in the interval of weekday/daytime/autumn. The effectiveness of the dynamic PTDF family selected is shown by comparing the estimated power flow using the dynamic PTDF matrices, a fixed PTDF matrix and the real power flow calculated in PSSE, given the economic dispatch run by Powrsym4 simulation without PTDF logic. As can be seen from Figure 4.18 and Figure 4.19, power flow estimation accuracy is improved substantially by adopting the dynamic PTDF. The peak value of power flow estimation error in the given case is reduced by 67% compared with using only one PTDF matrix. The average error of the estimated flow reduced from 16.9% with one fixed PTDF matrix to 3.7% when we adopt the PTDF family.

The implementation scheme of the methodology is illustrated in Figure 4.20.

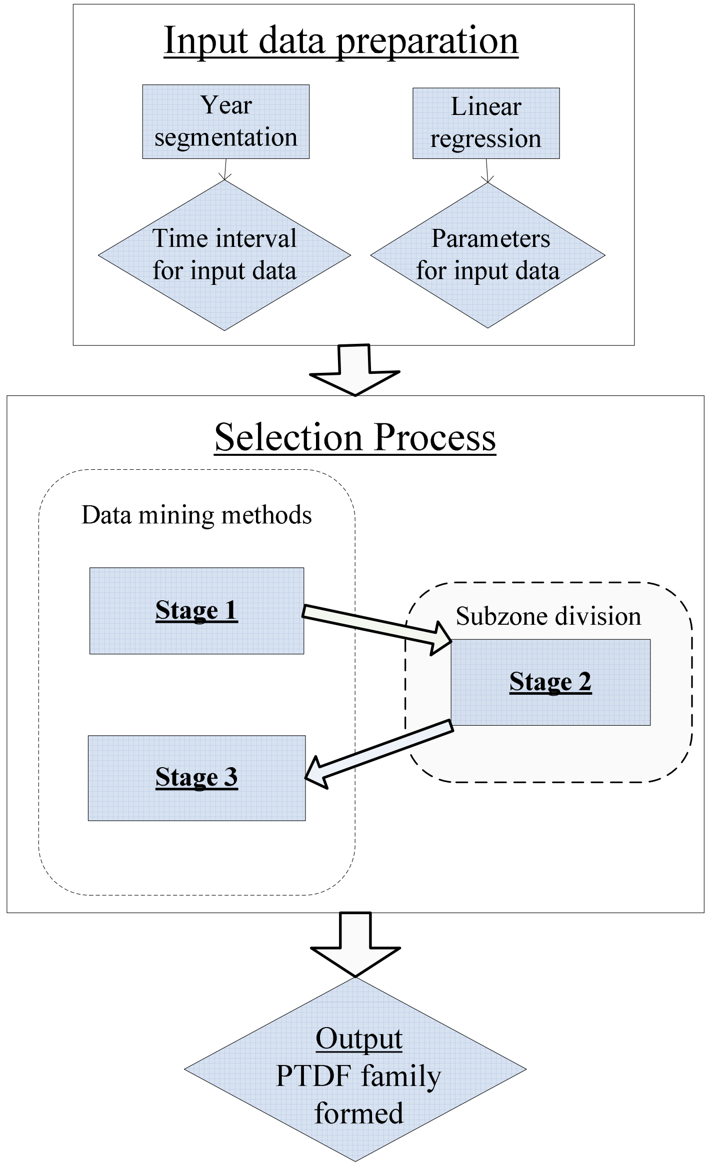

Figure 4.20 Procedures for dynamic PTDF family formation and representative PTDF selection

First, mathematical tool such as clustering method and silhouette value and is brought to the application of electricity market model to explore the typical scenarios. The k-medoids method finds the typical scenarios and the silhouette value indicates the optimal number of typical scenarios.

In order to gain more precise information for the input data set, the influence factors of PTDF matrix discussed in chapter 3 are taken into account in the implementation of mathematical tools. The input data is initially segmented to weekday/weekend, daytime/morning/evening and spring/summer/autumn/winter groups to consider the temporal feature. And subzone is divided by grouping buses with similar PTDFs in one area after the first round of k-medoids and silhouette value analysis (stage1) to add the special feature.

By defining the distances, we assess the similarity between the input data of a random hour to the typical scenario hours. The criteria to select a representative scenario for the random hour from the typical scenarios are to find the one with least distance. The distance also reflects the outliers of input data which we make use of. The typical scenarios selected from the outlier data set can represent the cases with more extreme input data well and are incorporated into the overall PTDF family.

It should be noticed in the study case, a difficult case is used to demonstrate all the steps where the outlier level in the input data is high. For the time intervals with input dataset where the data objects are close to each other, it is not necessary to perform the complete three stages. Data process can exit after stage 2 if all distances obtained in this stage are below the threshold predefined.

Chapter 5 Economic Dispatch with Dynamic PTDF Logic

5.1 Introduction

In chapter 4, the dynamic PTDF family selection was discussed in detail and it was used to improve the accuracy of power flow estimation given the economic dispatch without implementing PTDF logic. In this chapter, Matlab model is built for incorporating the dynamic PTDF family into the unit commitment and economic dispatch decision. The UC-ED results and PSSE power flows are compared for cases without PTDF logic, with simple PTDF logic (only one fixed PTDF matrix) and with dynamic PTDF matrix.

The chapter is organized as follows. First the economic dispatch model in Matlab is discussed. The security constraint economic dispatch model adopts the classic PTDF (See chapter 3) to estimate inter area power flow. Then the dynamic PTDF logic is implemented in the economic dispatch model is and its influence on the economic dispatch and on the resulted power flow is investigated.

5.2 Economic Dispatch Model with classic PTDF in Matlab

5.2.1 Linear Programming

Linear programming in Matlab has the following form:

$$\min f^T X \tag{5.1}$$

s.t.

$$A X \leq b \tag{5.2}$$

$$A_{eq} X = b_{eq} \tag{5.3}$$

$$LB \leq X \leq UB \tag{5.4}$$

Where A and A_{eq} are matrices and f , X , b , b_{eq} , LB and UB are vectors. The objective of the linear programming is to find a vector X that minimizes the objective function 5.1

while the equality constraint in equation 5.3 and inequality constraint in 5.2 and 5.4 are satisfied [19].

The objective function, equations and inequations of power system economic dispatch problem is based on Ref [20].

The objective of economic dispatch is to minimize the total operational cost of all the committed units in the given time interval.

$$\min F_G = \sum_{i=1}^{N} F_{Gi}(P_{Gi}) \tag{5.5}$$

Where

F_G is the total cost of the whole system.

F_{Gi} is the operational cost of each unit i.

P_{Gi} is the production of unit i.

N is the number of unit in the system.

Constraints of power system include power balance constraints, generating capacity constraints and security constraint. Assume there is no loss in the network. The power balance constraint states that the sum of power outputs of all units should be equal to the total power demand in the system.

$$\sum_{i=1}^{N} P_{Gi} = P_D \tag{5.6}$$

Where

P_D is the total power demand.

The generation capacity constraint is that the power output of any committed unit should be higher than or equal to its minimum power level and should be lower than or equal to its maximum power level.

$$P_{Gi}^{\min} \le P_{Gi} \le P_{Gi}^{\max} \tag{5.7}$$

Where

$P_{Gi}^{\min}$ is the minimum power level of unit i.

$P_{Gi}^{\max}$ is the maximum power level of unit i.

The security constraint we take into account in the market model is that the power flow on inter area links should not exceed their capacity limits.

$$\left| P_f \right| \le P_f^{\max} \tag{5.8}$$

The operational cost $F_{Gi}(P_{Gi})$ is usually characterized by second-order polynomial function:

$$F_{Gi}(P_{Gi}) = a_i + b_i P_{gi} + c_i P_{gi}^2 \tag{5.9}$$

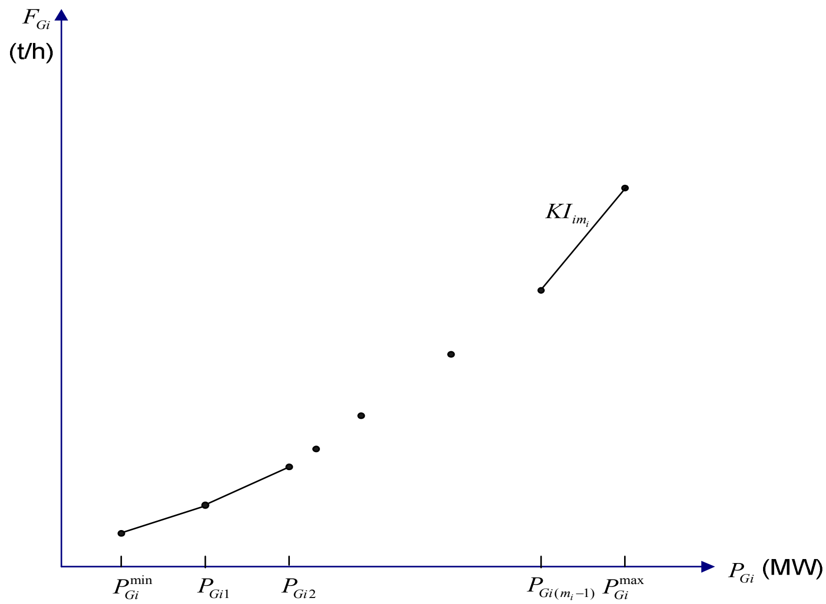

Figure 5.1 Operational cost of a thermal power plant unit

The power output range of each unit i is segmented to m_i pieces. Power output upper limits of the m_i pieces are $P_{Gi}^{\min}, P_{Gi1}, P_{Gi2}, \cdots, P_{Gi(m_i-1)}, P_{Gi}^{\max}$.The length of these pieces are $PI_{i1}, PI_{i2}, \cdots, PI_{im_i}$ and the average slopes are $KI_{i1}, KI_{i2}, \cdots KI_{im_i}$.

The operational cost can be linearized and expressed as:

$$F_{Gi} = F_{Gi}(P_{Gi}^{\min}) + PI_{i1} \times KI_{i1} + PI_{i2} \times KI_{i2} + \cdots + PI_{imi} \times KI_{im_i} \tag{5.10}$$

$$P_{Gi} = P_{Gi}^{\min} + PI_{i1} + PI_{i2} + \cdots + PI_{im} \tag{5.11}$$

The objective function can be written as:

$$\min F_G = \min\{\sum_{i=1}^{N}[F_{Gi}(P_{Gi}^{\min}) + PI_{i1} \times KI_{i1} + PI_{i2} \times KI_{i2} + \cdots + PI_{imi} \times KI_{im_i}]\} \quad (5.12)$$

$$\sum_{i=1}^{N}(P_{Gi}^{\min} + PI_{i1} + PI_{i2} + \cdots PI_{im_i}) = P_D \quad (5.13)$$

$$\begin{aligned} &0 \le PI_{i1} \le P_{Gi1} - P_{Gi}^{\min} \\ &0 \le PI_{i2} \le P_{Gi2} - P_{Gi1} \\ &\qquad \cdots \\ &0 \le PI_{im_i} \le P_{Gi}^{\max} - P_{Gi(m_i-1)} \\ &i = 1, 2, \cdots, N \end{aligned} \quad (5.14)$$

And the power flow on the inter area link can be expressed as

$$\vec{P}_f = PTDF^{zonal} * \Delta\vec{P}_B + \vec{P}_f^{\mathrm{Re}f} \quad (5.15)$$

$$\Delta\vec{P}_B = \vec{P}_G - \vec{P}_L \quad (5.16)$$

$\Delta\vec{P}_B$ is the vector of net production in each area. It is calculated by subtracting the vector of area load $\vec{P}_L$ from the area generation vector $\vec{P}_G$. The area generation is calculated by first summing up the power output pieces of units in one area, then adding up the minimum power level of the committed units in the area.

By comparing the objective function of economic dispatch problem 5.12 and constraints from 5.13 to 5.16 with the standard linear programming form in Matlab, we can compute the required matrices and vectors for linear programming to obtain the optimal power output in each piece of the committed plant. It is important to notice that unit commitment is needed as an input condition for the model.

5.3 Implementation of Dynamic PTDF Matrix on UC-ED Decision Making

First, the constraints in Table 5.1 are set in Powrsym4 without using PTDF logic and UC-ED is run.

Inter area link	Capacity limit
Link A1-A2	300
Link A1-A3	300
Link A2-A3	200

Table 5.1 Transmission link capacity limit

A random hour (hour 6561) in the interval of weekday/daytime/autumn is chosen for the following analysis. The input data of this hour is used for a whole week to obtain UC-ED so that the unit commitment output is specifically for this hour. The power flow reported in Powrsym and the real flow calculated in PSSE are shown in Figure 5.1 and Figure 5.2 respectively.

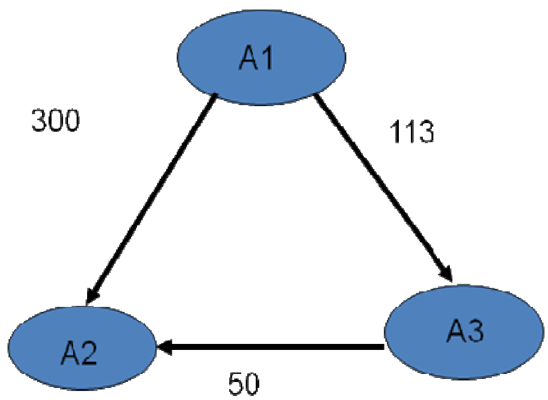

Figure 5.1 Power flow result in Powrsym4 without PTDF logic

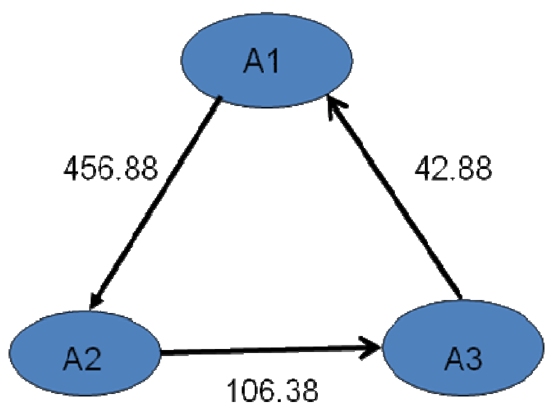

Figure 5.2 Real flow calculated by PSSE using the UCED result from Powrsym4

Next, the unit commitment and economic dispatch from Powrsym4 is used to compute distances between this hour and the typical scenario hours for choosing a typical scenario that can represent hour 6561 (See section 4.4). Scenario of typical hour 9 has the least distance with the scenario of this hour (hour 6561), determined by Powrsym4 without implementing PTDF at this stage.

The unit commitment decision from Powrsym4 for the previously selected random hour (hour 6561), and the PTDF matrix, net area balance and power flow of typical scenario 9 are used as input to obtain the economic dispatch in Matlab. In the same constraints as shown in Table 5.1, no feasible solution can be given. Th en the power flow constraint is removed in Matlab code with the input to get the economic dispatch in the situation without power flow constraint so that we can see which link is first congested. The resulted power flow estimation in Matlab is shown in Figure 5.3.

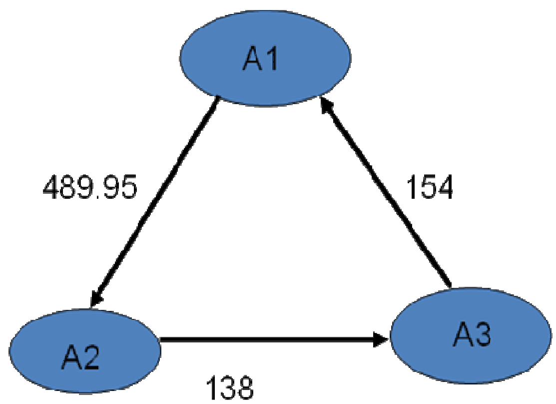

Figure 5.3 Power flows estimated by typical scenario 9 in Matlab without link capacity constraint

The real power flow calculated by PSSE with the UC-ED obtained in Matlab using typical scenario 9 in the situation without transmission link capacity limit is shown in Figure 5.4.

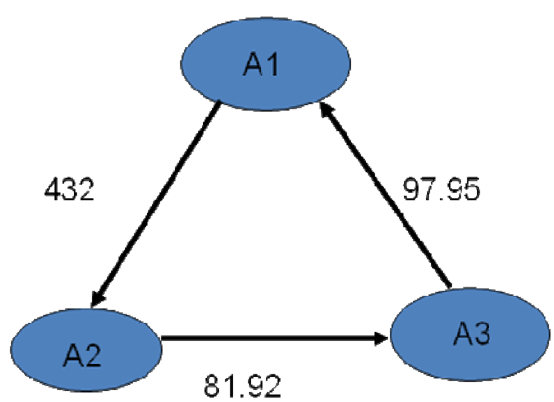

Figure 5.4 Power flows calculated by PSSE with the UC-ED obtained in Matlab using typical scenario 9 without transmission link capacity limit

Figure 5.3 demonstrates that the link from Area 1 to Area 2 is congested. The zonal PTDF of the typical scenario 9 is shown in Table 5.2 and the zonal transfer PTDF is calculated using equation 2.4 and shown in Table 5.3.

	A1	**A2**	**A3**
Link A1-A2	0.8171	0	0.6272
Link A1-A3	0.1828	0	-0.6273
Link A2-A3	-0.1828	0	-0.3727

Table 5.2 Zonal PTDF of typical scenario 9

	Transfer A1> A2	**Transfer A1 > A3**	**Transfer A2 > A3**	**Transfer A2 > A1**	**Transfer A3 > A1**	**Transfer A3 > A2**
LinkA1-A2	0.8171	0.1898	-0.6273	-0.8171	-0.1898	0.6273
LinkA1-A3	0.1829	0.8102	0.6273	-0.1829	-0.8102	-0.6273
LinkA2-A3	-0.1829	0.1898	0.3727	0.1829	-0.1898	-0.3727

Table 5.3 Transfer zonal PTDF of typical scenario 9

When we examine the transfer PTDF of typical scenario 9 in Table 5.3, it is easy to see from the first row of the transfer PTDF matrix that power transfer from A1 to A2 contributes most to the congestion. To alleviate the congestion on the link from Area1 to Area2, less transfer should be made from Area1 to Area2. There are several alternative ways to supply the load in Area2. Transfer can be made from other areas which have a smaller PTDF value or negative PTDF value on this link or the uncommitted plant can be started so less power import is required for Area2. The solution with the least cost is chosen (the system should be in balance for each solution).

The two committed plants in Area 2 produce at the maximum power level and there is only one gas engine plant (Plant 32) that is not committed. The plants that have lower operational cost in Area 3 such as wind power plant 41 and coal plant 38 have also dispatched to the maximum power level.

The possible options to solve the congestion are to commit power plant 32 in Area 2 or to commit the gas engine plant 33 or commit the CCGT plant 35 in Area3. Plant 32 has a lower operational cost than the plants in Area3. Moreover, the power production in Area 2 can better relieve the congestion. Power injection in Area 2 by 1MW to supply load in Area 2 can reduce 0.8171 MW power flow on the congested link. But for the same purpose of supplying load in Area2, by replacing power transfer of 1 MW from Area1 to Area2 with the same amount from Area3 to Area2, only 0.1898 MW flow can be reduced on the same congested link. Thus committing power plant 32 is the optimal solution.

The new economic dispatch result is obtained in Matlab with the constraints in Table 5.1. Estimated power flow using typical scenario 9 is shown in Figure 5.5.

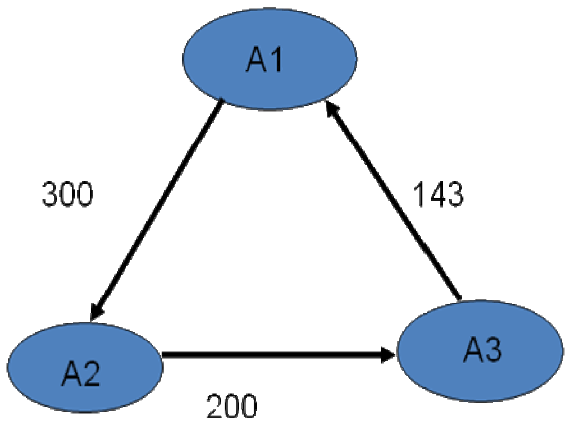

Figure 5.5 Power flow estimated in Matlab based on typical scenario 9 with plant 32 committed

And the real power flow calculated in PSSE is illustrated in Figure 3.6. The error of the estimated power flow is 20.4 MW.

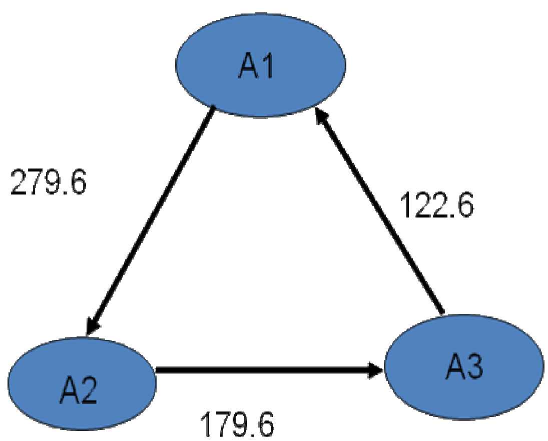

Figure 5.6 Power flow calculated by PSSE with the UCED obtained in Matlab using typical scenario 9 without transmission link capacity limit

Then distance is calculated between the current hour scenario with economic dispatch given by Matlab and the typical scenario 9 using equation 4.7 with variables shown in Table 5.4 and regression parameters in Table 4.1.

	WA1 (MW)	WA2 (MW)	WA3 (MW)	LA1 (MW)	LA2 (MW)	LA3 (MW)	NAB1 (MW)	NAB2 (MW)	SAB1 (MW)	SAB2 (MW)
Matlab ED	131	173	83	942	1623	1753	157	-100	726	-783
TS9	118	315	128	1009	1884	2362	666	-309	582	-939

Table 5.4 Matlab economic dispatch scenario and the typical scenario 9

The distance calculated is 1604MW. Assuming a predefined threshold of 1700MW for the New England system, the distance is in an acceptable range. If the distance was higher than the threshold, that would be an indicator that the estimated power flow might deviate largely from the real value. In this case, distances of the scenario based on the economic dispatch obtained from the initially chosen typical scenario with all typical scenarios need to calculated. And the typical scenario that has least distance is chosen as the new representative scenario to recalculate economic dispatch. The process iterates until the distance is no larger than the threshold.

In the following part, we use a fixed PTDF for economic dispatch instead of choosing from a dynamic PTDF family to obtain the economic dispatch scenario and power flow for comparison. The scenario and PTDF of the first hour of the year is used as the reference for this hour to obtain the economic dispatch, there is no feasible solution reported by the Matlab. After the transmission link capacity limit is removed, we can see the link from Area1 to Area2 is congested. So power plant 32 is committed for the reasons as we discussed in the case of typical scenario 9.

Figure 5.7 shows the power flow estimated in Matlab with plant 32 committed and by using the PTDF matrix of the first hour of the study year.

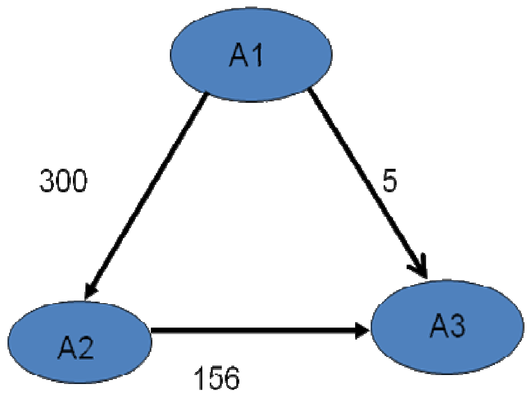

Figure 5.7 Power flow estimated in Matlab using the fixed PTDF

Figure 5.8 shows the power flow calculated in PSSE based on scenario of the first hour of the study year with plant 32 committed. The error of estimated flow is 39 MW, which doubles compared with the value of adopting typical scenario 9.

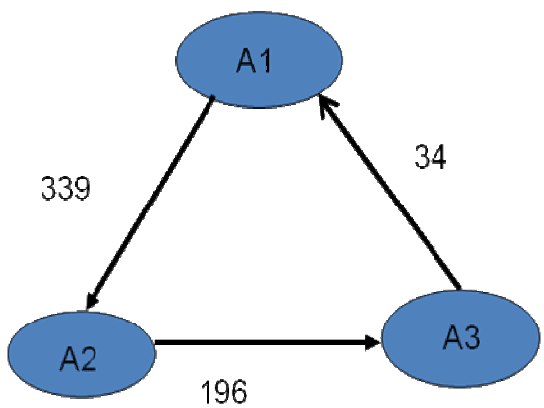

Figure 5.8 Power flow calculated in PSSE using fix PTDF

The economic dispatch result by Powrsym and Matlab are shown in Table 5.4.

ED decision made by / Power plant number	Powrsym without implementing PTDF logic (MW)	Matlab typical scenario 9 with unlimited link capacity (MW)	Matlab typical scenario 9 with plant 32 committed (MW)	Matlab fix PTDF plant 32 committed (MW)
A1 Plant30	260	180	150	150
A2 Plant 31	173	173	173	173
A2 Plant 32	0	0	251	208
A3 Plant 33	0	0	0	0
A3 Plant 34	217	291	218	180
A3 Plant 35	0	0	0	0
A3 Plant 36	339	345	345	278
A1 Plant 37	965	966	818	966
A3 Plant 38	1050	1050	1050	1050
A2 Plant 39	1099	1099	1099	1099
A1 Plant 40	131	131	131	131
A3 Plant 41	83	83	83	83

Table 5.4 Different economic dispatch results

5.4 Conclusion

In this chapter, Economic dispatch model is constructed in Matlab and the algorithm is discussed in detail. This model considers the power flow constraints with classic PTDF (See Chapter 3), which makes it applicable to simulate the result of implementing dynamic PTDF family.

The case without incorporating PTDF logic in market model is constructed and UC-ED and run by Powrsym4. A random hour is chosen for the analysis. For this hour, a significant error of the reported power flow on inter area link in Powrsym4 can be seen.

A comparison is made between adopting the typical scenario from dynamic PTDF family and the scenario of the first hour of the study year in the economic dispatch model. As demonstrated in section 5.3, by using the typical scenario, the production from the area that contribute most to the congested link will be reduced compared with using a fixed scenario (first hour of the year in the example). Moreover, the estimated power flow from the economic dispatch result based on typical scenario is closer to the real power flow than the case with a fixed scenario. As can be seen from Figure 5.5 to Figure 5.8, the absolute error of using typical scenario 9 is only half compared with using the scenario from the first hour of the year.

The proposed implementation scheme of the dynamic PTDF family for the UC-ED decision making is illustrated in Figure 5.9.

In this chapter, only one hour is simulated in the Matlab model due to the time limit. More simulation should be performed in a longer time series for getting better insights on the implementation of dynamic PTDF logic.

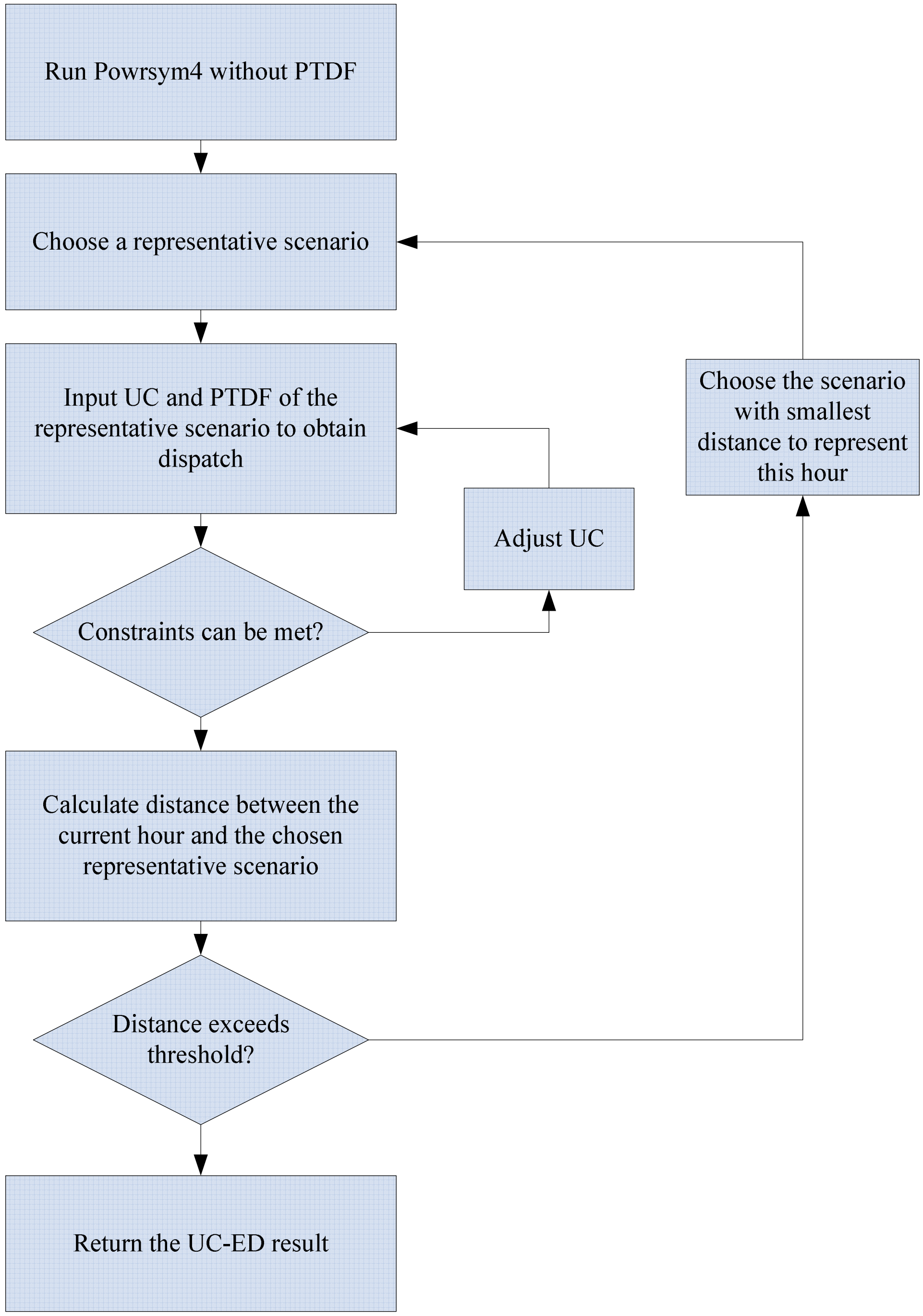

Figure 5.9 implementation scheme of the dynamic PTDF family for the UC-ED

Chapter 6 Conclusions and Recommendations

The main motivation of this research is a more accurate representation of the transmission grid within the market simulations. Only using NTC values without considering the inter dependencies between different links can cause congestion in the real grid. That is why the PTDF representation of the cross border links is a step forward. The common practice for implementing PTDF currently by TSO is to use a fixed PTDF matrix for all the hours with the purpose of power flow estimation after UC-ED result is obtained from market model simulation (See chapter 3). As demonstrated in the master thesis, the fixed PTDF matrix still brings a substantial error to estimate power flow, thus more representative PTDF matrices have to be used for different hours.

This master thesis aims at developing the methodology to form dynamic PTDF families, for the implementation into the market model. The main procedures to create a dynamic PTDF family are developed and presented with a study case (See Chapter 4). By fitting data from the market model into the data mining tools, typical scenarios of the time intervals are selected. Furthermore, by dividing subareas according to nodal PTDF values, the resulted estimation power flow is getting closer to the real flow, as depicted in section 4.4 for the study case. In addition, the outliers of the initial input data are located and more typical scenarios are selected from this outlier data group. The effectiveness of the dynamic PTDF family is demonstrated by comparing the estimated power flow using the dynamic PTDF matrices, a fixed PTDF matrix and the real power flow calculated in PSSE, given the same economic dispatch run by Powrsym4 simulation without PTDF logic. The results reveal that with the dynamic PTDF family adopted, the power flow estimation error is significantly reduced in comparison with using a fixed PTDF matrix. In the study case, the peak value of the estimation error is reduced by 67%. The average error of the estimated power flow reduced from 16.9% with one fixed PTDF matrix to 3.7% when we adopt the PTDF family.

In the last part of the thesis, the implementation of dynamic PTDF matrix in the market model is simulated with a Matlab economic dispatch model for one hour. The influence of incorporating dynamic PTDF matrix into the UC-ED decision is investigated. The UC-ED results and PSSE power flows are compared for cases without PTDF logic, with simple PTDF logic (only one fixed PTDF matrix) and with dynamic PTDF matrix. At last, an implementation scheme of the dynamic PTDF family for the UC-ED decision making is proposed. However, due to the time limit, only one hour was simulated in this master thesis. More simulations should be performed over a longer time horizon for future work. The implementation of the dynamic PTDF logic into a chronological market model such as Powrsym4 should be investigated.

A next step would be to implement and adapt the method for a larger and more complex power system. For the application in other systems, it should be noticed that the regression coefficients used for distance calculation and the distance threshold are specific for each system. The regression coefficients should be recalculated, and some tests should be performed to determine a distance threshold for the larger system to control the power flow estimation error in an acceptable range while the number of dynamic PTDF matrices is reasonable.

Chapter 7 References

[1] P.Schavemaker and L.van der Sluis, Electrical Power System Essentials, 2008

[2] B.C. Ummels, Power System Operation with Large-Scale Wind Power in Liberalised Environments, 2009.

[3] Trevino, Luis, Liberalization of the Electricity Market in Europe: An overview of the electricity technology and the market place, 2008.

[4] Directive 2003/54/EC of the European Parliament and of the Council of 26 June 2003 concerning common rules for the internal market in electricity, available at http://europa.eu/legislation_summaries/energy/internal_energy_market/l27005_en.htm

[5] ETSO-EuroPex, Flow-based Market Coupling A Joint ETSO-EuroPEX Proposal for Cross-Border Congestion Management and Integration of Electricity Markets in Europe,2004.

[6]] Regulation 1228/2003, available at http://www.energy-community.org/portal/page/portal/ENC_HOME/AREAS_OF_WORK/ELECTRICITY/Regional_Market/Regulation_1228

[7] Andreas Ehrenmann, Yves Smeers, 2004: Inefficiencies in European Congestion Management. Available on: http://www.hks.harvard.edu/hepg/Papers/Ehrenmann.Smeers_Inefficiences.European.Congestion.Mgmt.pdf.

[8] JÜrgen Apfelbeck, Jean-Baptiste Bart, et al. EMELIE-NET, a game theoretic approach including PTDF, 2005.

[9] Purchala, K., E. Haesen, et al. (2005). Zonal network model of European interconnected electricity network, CIGRE/IEEE PES, 2005, International Symposium.

[10] OSA, DESCRIPTION OF THE *Powr*Sym3™ COMPUTER MODEL

[11] M. Korpås, L. Warland, Further Developing Europe's Power Market for Large Scale Integration of Wind Power,

[12] John J. Grainger, William D. Stevenson, Power system analysis,1994

[13] Jamie Weber, Efficient Available Transfer Capability Analysis Using Linear Methods, November 7, 2000. PSERC Internet Seminar.

[14] Optimization of Transmission Systems by use of Phase Shifting Transformers, Jody Verboomen,2008.

[15] C.Duthaler, M.Kurzidem, M.Emery, G.Andersson Analysis of the Use of PTDF in the UCTE Transmission Grid, 16th Power Systems Computation Conference (PSCC), Glasgow, 14th−18th July, 2008.

[16] Ercot, Ercot Protocols Section 7: Congestion Management, July 1, 2010

[17] M. Vukasovic and S. Skuletic,Implementation of different methods for PTDF matrix calculation in flow-based coordinated auction, POWERENG 2007, April 12-14, 2007, Setubal, Portugal.

[18] Hung-po Chao and Stephen Peck, A market mechanism for electric power transmission, Journal of Regulatory Economics, Vol. 10, No. 1, 1996, pp. 25-59.

[19] Thomas S. Ferguson, LINEAR PROGRAMMING A Concise Introduction.Available on http://www.usna.edu/Users/weapsys/avramov/Compressed%20sensing%20tutorial/LP.pdf

[20] Jichun Liu, Power system economic dispatch optimization theory and its application, 2010

[21] Gross, Jürge, Linear Regression, 2003

[22] Jiawei Han, Micheline Kamber, Data Mining Concepts and Techniques, 2006

[23] Dan Nettleton, Cluster Analysis for Microarray Data, Seventh International Long Oligonucleotide Microarray WorkshopTucson, Arizona January 7-12, 2007 http://www.cals.arizona.edu/microarray/Jan07Workshop/lectures/Lecture%2010.pdf

[24] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis," *J Comput. Appl. Math.*, 20, 1987, 53-65

Appendix

Mathematical Methods

Linear Regression

Linear Models

Linear regression evaluates the linear relationship between independent variables X and dependent variable y.

The linear regression model [21] is described by equation:

$$y = X\beta + \varepsilon \tag{1}$$

where

$$y = \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{pmatrix},\; X = \begin{pmatrix} x_{1,1} \cdots x_{1,p} \\ \vdots \qquad \vdots \\ \vdots \qquad \vdots \\ x_{n,1} \cdots x_{n,p} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \tag{2}$$

A linear relationship is expressed between a variable y and variables $x_1, \cdots, x_p$ with small error collection represented by non observable variable ε. The variable y is called dependent variable and variables $x_1, \cdots x_p$ are called independent variables. The $n \times 1$ vector y consists of sample variables, while X is a non-stochastic $n \times p$ matrix comprises of independent variables with $p < n$. The $n \times 1$ vector ε is the vector of non-observable random variables which fulfills $E(\varepsilon) = 0$ and $Cov(\varepsilon) = 0$ with $\sigma^2 = 0$.

Often linear regression model is expressed with intercept. That is, one of the independent variable is equal to 1, so that one column of matrix X is n ones. The linear regression model is written as

$$y = 1_n \mu + X\beta + \varepsilon \tag{3}$$

μ is the intercept parameter which represents the constant in linear relationship so that matrix X contains only 'real' independent variables.

Proceeding with Linear Models

To investigate linear relationship between variables of interest, the following three stages are performed:

Stage 1: Model Building

In this stage, the relationship that we wish to investigate must be specified. Some software packages can be used to carry out the variable elimination process to determine whether a variable should be adopted in the model. Some plots can be used to find extreme values which worth eliminating.

Stage 2: Inference

After the model is established, the relationship between the dependent variable and independent variable can be determined. In this stage, the unknown parameters can be obtained.

Stage 3: Prediction

The final goal of the analysis is usually to predict y with some given values ($x_{m,1},\cdots,x_{m,p}$). One can expect better prediction of y with more adequate model and more precise estimates obtained in stage 2.

Clustering Method

Clustering [22] is the process of partitioning objects into classes that consist of similar objects. A cluster is a group of data objects that have high similarity within the same cluster but are dissimilar to objects in other clusters.

Categorization of Clustering Methods

The major clustering methods [22] are categorized:

Partitioning methods: With a given dataset of n objects and the desired number of cluster k, a partitioning method groups the data into k clusters. It uses an iterative relocation technique that tries to improve the quality of the partition by moving objects from one group to another. The most commonly used partitioning methods are k-means algorithm and k-medoids algorithm. The clusters are formed to minimize the dissimilarity function within a cluster, which is always measured by distance. In k-means method, the distance is measured between objects and the cluster mean, while in k-medoids method it is measured between objects and the representative object.

Hierarchical methods: The hierarchical methods produce nested sequence clusters. There are two types of hierarchical methods. The agglomerative hierarchical clustering uses the bottom up approach by first placing each cluster in its own cluster and then merges them into larger cluster, until all the objects belong to a single cluster or certain termination condition is met. The divisive hierarchical clustering uses the top-down strategy that processes the data in a reversed manner. Initially all the objects are in one cluster and are subdivided into smaller clusters until each object becomes a cluster or certain termination condition is satisfied. The disadvantage of hierarchical methods is that the quality of the clustering suffers from its inability to adjust once a step (merge or split) is undone.

Density-based Methods: Unlike partitioning methods, which group data based on the distance between objects and find spherical-shaped clusters, the density-based methods construct clusters based on the density (number of objects or data points). For a given cluster, if the density of neighbor cluster exceeds threshold, it will continue to grow. Density-based methods are applied to discover clusters with arbitrary shape.

Grid-based Methods: The grid-based method quantizes the data space into cells that form a grid structure.

Model-based Methods: Model-based methods hypothesize mathematical model for every cluster and find the best fitted data to the given model.

Since the data from the market model is numerical and we are only interested to obtain the centers of cluster to represent the typical scenarios for market simulation, the partitioning method is chosen.

Requirements for Clustering Method

Requirements for clustering method [22] that are related with this master thesis are listed below:

Scalability: Many clustering algorithms can be applied well on small data sets which have fewer than several hundred data objects. Clustering on a large data set may lead to biased result.

Discovery of clusters with arbitrary shape: Many clustering method determine cluster based on Euclidean or Manhattan distance.

Requirements to determine input parameters: Many clustering methods require users to input parameters such as the number of clusters desired.

Ability to deal with noisy data: Some clustering methods are sensitive to outliers in the data set and lead to poor clustering result.

High dimensionality: Many clustering methods are good at handling low-dimensional data, but clustering data with high dimensions is challenging.

Euclidean Distance

The dissimilarity between objects is measured on the distance between each pair of objects. The distance used in this master thesis is Euclidean distance:

$$d(i,j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \cdots (x_{in} - x_{jn})^2} \qquad (4)$$

Where

$i = (x_{i1}, x_{i2}, \cdots x_{in})$ are n-dimensional data objects.

$j = (x_{j1}, x_{j2}, \cdots x_{jn})$ are n-dimensional data objects.

And the Euclidean distance satisfies the following mathematical requirements:

1. $d(i, j) \geq 0$: Distance is nonnegative
2. $d(i,i) = 0$: For any object, the distance to itself is 0.
3. $d(i, j) = d(j,i)$: Distance is symmetric.
4. $d(i, j) \leq d(i,h) + d(h, j)$: Distance between object i and object j is no more than that of making a detour over another object h.

Partitioning Methods

The k-means Method

The k-means algorithm [22] partitions the given data set with n objects into k clusters. Each cluster is represented by the mean value of the objects that are assigned to the cluster. The k-means algorithm requires the number of clusters k as input.

First, it selects k of the objects randomly, to be the initial cluster center or cluster mean. Then for each of the rest objects, the distances between the object and the k cluster means are calculated. The object is assigned to the cluster whose center has the smallest distance from it. After all the objects are distributed to a cluster, the new means are recalculated based on the current objects in the cluster. Using the updated cluster means, each object is assigned to the cluster which has the nearest center from it. The process iterates until no redistribution of objects occurs.

But the k-means method suffers from the fact that the mean value is sensitive to outliers so that an object with extreme values may distort the distribution of data substantially.

The k-medoids Method

The k-medoids method [22] picks actual object to represent the clusters. Similar to k-means, the number of clusters k is required as input.

An example [23] is used to illustrate the steps of k-medoids:

Step1: The initial k cluster centers or cluster medoids are chosen arbitrarily.

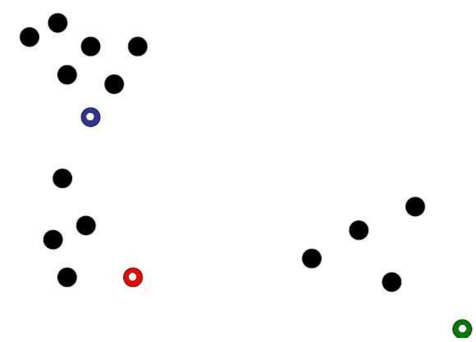

Figure 2 Choosing initial cluster centers arbitrarily [23]

Step2: For each of the remaining objects, the distances between the object and the k medoids are calculated and the object is assigned to the cluster which has the nearest medoid.

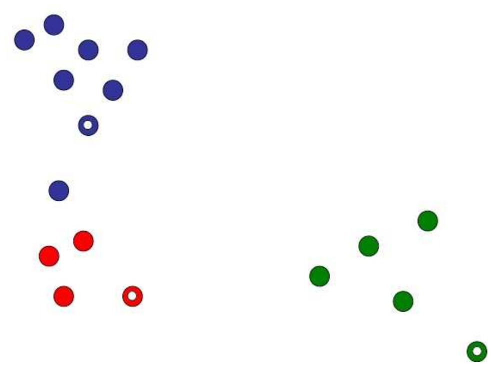

Figure 3 Assigning objects to clusters [23]

Step3: After all the objects are distributed and clusters are formed, the new medoid of each cluster is calculated by finding the one with least average distance to the rest of objects in the cluster.

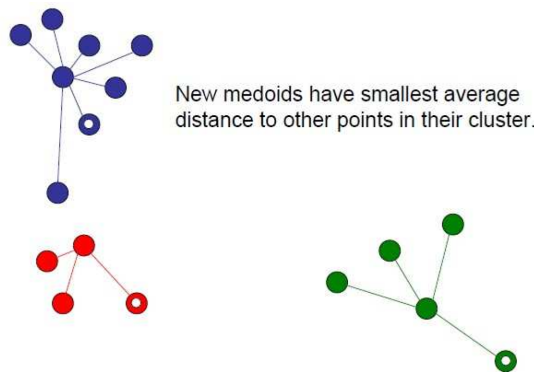

Figure 4 Forming new cluster centers [23]

Step4: Using the updated cluster medoids, each object is assigned to the cluster which has the nearest medoid. The process iterates until no redistribution of objects occurs.

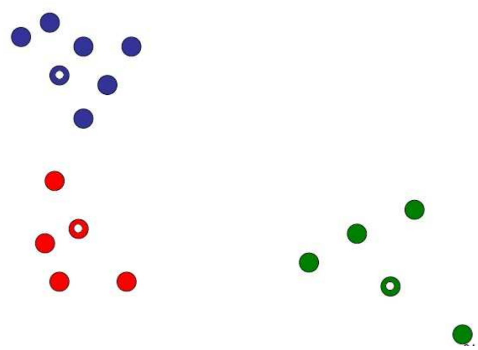

Figure 5 Assigning objects to new clusters [23]

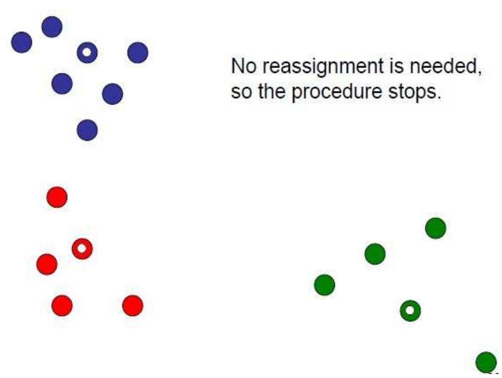

Figure 6 Forming the final clustering result [23]

To conclude, k-medoids method offers the advantage to reduce the influence of outliers. It is a more robust method compared with k-means method, hence the k-medoids method is chosen for the application of selecting typical scenarios in this master thesis.

Optimal Number of Clusters

Silhouette Value

Silhouette value [24] assesses how well an object i fits in the current cluster it is assigned to.

The calculation of silhouette value proceeds as follows. For any object i in cluster A, we can compute the average distance of all lines connected with i within i:

$$a(i) = \text{average dissimilarity of i to the rest of objects in A}$$

And for any other cluster, cluster C for example, the average distance between i and cluster C is:

$$d(i, C) = \text{average dissimilarity of I to all objects of C}$$

After computing $d(i, C)$ for all the clusters which are different from A, the one with the shortest distance is selected and denoted by $b(i)$.

$$b(i) = \underset{C \neq A}{\text{minimum}}\ d(i, C)$$

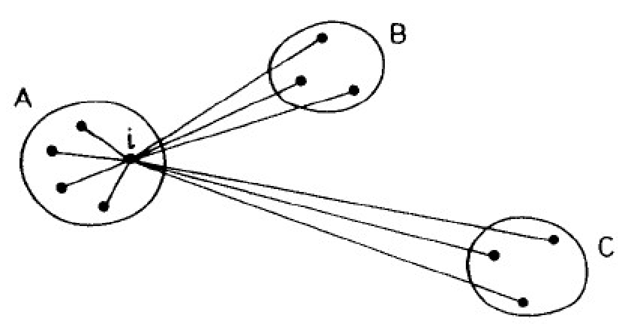

Figure 7 Illustration of the elements related in computation of s(i) [24]

The cluster B that has the shortest distance with i is called the neighbor of object i. And the silhouette value $s(i)$ is written as:

$$s(i) = \begin{cases} 1 - a(i)/b(i) & if\ a(i) < b(i), \\ 0 & if\ a(i) = b(i), \\ b(i)/a(i) & if\ a(i) > b(i). \end{cases} \quad (5)$$

It can be expressed in the form:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (6)$$

In the situation when there is only one object i in cluster A, we have $a(i) = 0$ and the silhouette value is set to zero.

If silhouette value is near 1, the dissimilarity within cluster A is far smaller than the dissimilarity between object i and neighbor cluster B, so i lies much closer to other objects in cluster A than to objects in cluster B. We can conclude that i is assigned to the appropriate cluster and or say object i is ‘well-clustered’.

If silhouette value is near -1, the dissimilarity between object i and cluster B is far smaller than the dissimilarity within cluster A, so object i is much closer to objects in cluster B than the other objects in cluster A. In other words, object i is better assigned to cluster B than to cluster A.

To sum up, silhouette value indicates if a single object is well clustered. When we consider all the objects in the data set, the *average silhouette value* is used to evaluate the overall quality of clustering.

Steps for Finding the Optimal Number of Clusters

The steps to attain the overall average silhouette value with a given data set and the number of cluster k is depicted in the flow chart below.

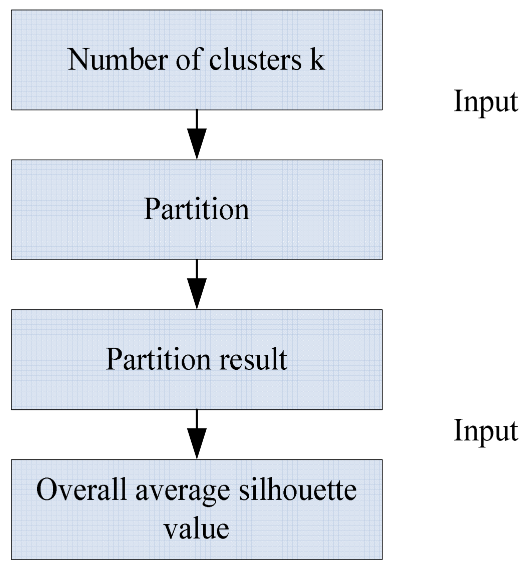

Figure 8 Steps to obtain the average silhouette value

Thus the overall average silhouette value is used to find the best number of clusters k by selecting the number of clusters with the highest value.