DISSERTATION

COMPUTATION OF LOOP FLOWS IN ELECTRIC GRIDS WITH

HIGH WIND ENERGY PENETRATION

Submitted by

Manish Mohanpurkar

Department of Electrical and Computer Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2013

Doctoral Committee:

    Advisor: Siddharth Suryanarayanan

    Liuqing Yang
    Peter M. Young
    Daniel J. Zimmerle

ABSTRACT

COMPUTATION OF LOOP FLOWS IN ELECTRIC GRIDS WITH

HIGH WIND ENERGY PENETRATION

In a deregulated electricity market, the financial transmission rights (FTRs) and the bid-sell principle for energy trades are used to determine the expected power flows on transmission lines. Expected power flows are calculated by applying the superposition theorem on the approved electronic tags (e-tags). Multiple parallel paths in interconnected networks lead to division of power flows determined by the impedances of the parallel paths and the physical laws of electricity. The actual power flows in the network do not conform to the market expectations leading to unscheduled flows (USF) on transmission lines. USF have historically been estimated and accommodated deterministically for a given set of e-tags. However, wide-area interconnections experience variability and uncertainty due to a significant penetration of wind energy connected at the transmission level, thus imparting a stochastic nature to USF.

A linear model, from the literature, has been adopted to model USF using a mathematical artifact called '*minor loop flows*'. This research develops an automated framework that provides accurate estimates of loop flows suitable for both market and network level accommodation of variable USF. This generic framework will be applicable to any power transmission network with intermittent energy resources.

A loop detection algorithm (LDA) based on graph theory is proposed to detect loops in a transmission network of any size. The LDA is formulated as a modification of the A-star (A*) algorithm, the lowest ancestor theorem, and Dijkstra's algorithm. The LDA has an order of complexity of $\mathcal{v}^2$, where $\mathcal{v}$ is the total number of vertices or buses in the network under

consideration. An application of a geographical information systems (GIS) technique has been established to obtain the transmission line layouts. The outcome of the LDA (i.e., minor loops) and line layouts (i.e., *azimuth*) are processed to compute the incidence matrix of the estimator. The variability due to the penetration of wind energy is accounted in the proposed framework using the probabilistic load flow analysis based on Monte Carlo simulations. Three techniques - ordinary least squares (OLS), analytic ridge regression (RR), and robust regression (M-estimators) - are used to estimate minor loop flows. The estimation techniques adhere to the auto-correction of the quality of estimates in case of ill-conditioning of the incidence matrix. Accuracy of loop flow estimates is highly significant, as they may be used for assigning economic responsibility of USF in electricity markets.

Wind power generation companies (WGENCOs) employ forecasting models to participate in the primary electricity markets. Forecasting models used to predict the output of wind power plants are inherently erroneous and hence, their impacts on USF are studied. The impact of forecasting errors associated with the output of wind plants is investigated using the concept of prediction intervals rather than point accurate forecasts. Loop flow estimates corresponding to the prediction intervals of power output of wind power plants are computed to provide statistical bounds.

The proposed framework is tested on the IEEE 14-bus and the IEEE 30-bus standard test systems with suitable modifications to represent wind energy penetration. Accurate loops are detected for the aforementioned test systems using the LDA.

Thus, an automated and generic computation of loop flows is proposed along with a step-wise demonstration on IEEE test systems is provided. Future work and concluding remarks summarize the research work in this dissertation.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

NOMENCLATURE

$*$	"Any" index. For example, e_{i*} indicates the set of edges between vertex i and any other vertex in G
$\hat{x}_{ols}$	Loop flow estimate using OLS
$\hat{x}_{rr}$	Loop flow estimate using analytic RR
$\frown$	Carat indicates a composite edge representing a path through multiple vertices
u_{ij}	Unidirectional edge from v_i to v_j
σ^2_{ols}	Variance of loop flows obtained using OLS
σ^2_{per}	Variance of loop flows obtained using proposed estimator
σ^2_{rr}	Variance of loop flows obtained using analytic
Δ	Bus angle
Π	Set of all graphs which are finite, connected, multi-planar, bidirectional, and sparse
$\mathcal{D}$	Degree of a vertex/graph calculated using the weight of 0.5 for a unidirectional edge
$\mathcal{O}$	Order of complexity
$adj_{u/b}[]$	Adjacency function i.e., list of adjacent vertices for the vertex listed within []; the subscripts u and b imply unidirectional and bidirectional
AEP	American Electric Power
AMI	Automated Metering Infrastructure
BPA	Bonneville Power Administration
C	Connectivity matrix of the graph G.
CDF	Cumulative distribution function
DDMS	Device Data Management System
DFS	Depth First Search
DISCO	Distribution Company
DoE	Department of Energy
EMS	Energy Management Systems
e-tags	Electronic tags

FCITC	First Contingency Incremental Transfer Capability
FERC	Federal Electricity Regulatory Commission
FTR	Financial Transmission Rights
G	Graph with known connectivity
GAPP	General Agreement on Parallel Paths
GENCO	Generating Company
GIS	Geographical Information Systems
H	Incidence or System Matrix
IEEE	Institute of Electrical and Electronics Engineers
IPF	Interface Participation Factor
ISO	Independent System Operator
K	Successive summation of weights of vertices, integer value.
KS	Kolmogorov Smirnov Test
l	Complexity of connection of a vertex v or alternatively the degree of the vertex
LDA	Loop Detection Algorithm
Level	Levels of the graph G registered during the wave search (two variables used for simplicity)
LMP	Locational Marginal Price
LODF	Line Outage Distribution Factor
$loop$	Loop array that consists of an ordered set of vertices.
MAE	Mean absolute error
MC	Monte Carlo simulations
MMWG	Multi-regional Modeling Working Group
m	Total number of wave levels in the graph G traversed by a wave front
n	Total no. of transmission lines in the network
NERC	North American Electricity Reliability Council

NYISO   New York Independent System Operator

OLS   Ordinary Least Squares

OPF   Optimal Power Flow

p Total no. of loops selected in the network

P   Real or Active power

PJM   Pennsylvania - New Jersey – Maryland Interconnection

POD   Point of Delivery

POR   Point of Receipt

PST   Phase Shifting Transformers

PTDF   Power Transmission Distribution Factor

Q   Reactive power

QCD   Qualified Control Devices

RR   Ridge Regression

RTEP   Regional Transmission Expansion Plan

S Storage bounds

SMD   Standard Market Design

SCUC   Security Constrained Unit Commitment

SimAuto   Simulation Automation Server

TRANSCO   Transmission Company

Traversed   Set of vertices visited by the imaginary wave in the graph G

V   Total no. of vertices in the graph G

V_{bus} Voltage magnitude of a bus bar

v_i Vertices of the graph G

V_{in} Inward wave direction chosen with respect to the $V_{start.}$

V_{start} Starting vertex for wave detection of loops

W_{Level}	Set of vertices traversed in the wave level
WECC	Western Electrical Coordinating Council
WECS	Wind Electric Conversion Systems
WGENCO	Wind Energy Generation Companies
x	Vector of loop flows
Y_{bus}	Admittance matrix of the power system network i.e., input of the algorithm.
z	Vector of USF

CHAPTER 1

INTRODUCTION

1.1   Motivation

The primary motivation of the proposed research is the quantification and impact analysis of variable unscheduled flows (USF) on account of the significant wind energy penetration expected in modern power systems. USF have been existent since the advent of interconnected power systems but gained critical importance after the deregulation of electricity industry. The nature of USF prior to the penetration of wind and solar powered plants was deterministic, since only conventional resources characterized the generation profile of the network. The nature of USF is determined by the actual power flows on transmission lines. For a strictly conventional generation profile, the sources of uncertainty associated with USF are forced generator and line outages and variability of loads. Hence, a snapshot of USF for a given contract period is considered sufficient for accommodation of USF at either the market or the operational level.

With the introduction and continued increase of large scale interconnections of wind power plants with the grids, the power flows on transmission lines possess a significant level of variability and uncertainty [1]. This imparts a stochastic component to the USF, as they are defined as the deviation from market expectations of power flows [2]. From the perspective of wind power plants, the level of variability and uncertainty stems from the wind regime and the errors associated with forecasting models used to predict the output of wind plants. To participate in the primary markets, wind power generation companies (WGENCO) have to commit a fixed value of power for the given contract period i.e., emulate a constant power source [3]. If the wind power plant generated the power equal to the committed value for a contract period then the nature of unscheduled flows will resort to being deterministic. In modern

1

electricity markets, wind power contracts may be treated as firm or non-firm depending on network characteristics. For example, in the Bonneville Power Administration's (BPA) balancing areas, energy from variable generation sources is tagged as 'firm' and hence are subject to stringent contractual obligations [4]. The unit commitments made by wind power plants are based on forecasting models which are inherently subjected to errors. Thus, the actual power flows when wind plants are connected in the network will inherit the stochasticity of the resource and also forecasting errors. Additionally, WGENCOs performance as either an importer during deficit generation or exporter during surplus generation from the secondary market also alters the USF scenario.

The proposed framework adopts a linear estimator that accommodates variable USF using a mathematical artifact called *minor loop flows* [2]. Deterministic USF are suspected to cause complications in cleared trades, the available transmission capacity, congestion of transmission lines, uncompensated use of transmission resources, and schedule curtailment [2]. Investigating the impacts of variable USF on the operation of the market and network operation is an additional motivation. The technical accommodation of USF by deploying qualified control devices at every bus of the network is cost prohibitive [2]. The rate and monetary charge of control power flow by qualified control devices is a function of severity of congestion on the regulated transmission line. This may grant an opportunity to the owner of a strategically located qualified control device to charge services at higher rates leading to a monopolistic situation. The aforementioned impacts and technical solution of using qualified control devices lead to a sub-optimal operation of the electricity market and hence is undesirable.

According to an executive summary by the General Agreement on Parallel Paths (GAPP) approximately 69% of the transmission revenue is wrongly allocated in bulk interconnections

[5]. A practical case of USF around Lake Erie with scheduled trades in the New York Independent System Operator (NYISO) reported a 70% adjustment charge in transmission revenues to settle inconsistencies in power flows [6]. The magnitudes of misallocated transmission charges signify the criticality of a financial mechanism that will assist in the appropriate allocation of transmission charges on account of variable USF in the networks.

USFs are one of the crucial seams issues that Independent System Operators (ISOs) have to deal with on a daily basis. They are not limited to any particular area, control, zones, etc. and are prevalent over the entire transmission network. USF may pose as a significant challenge both on network and market operational level for the ISOs. Managing USF needs a multitude of efforts on system expansion planning, inter-regional market coordination, advanced market mechanisms, strategic governance, and schedule optimization [7]–[10] The research presented in this dissertation was pursued as a part of the Regional Transmission Expansion Plan (RTEP) of the Western Electricity Coordinating Council (WECC). The objective of the RTEP is to develop a set of feasible network expansions for the Western interconnection of the United States. Constraints on expansion are to serve the expected increase in load; low cost of implementation; and, enabling the future plans of wind energy expansion intended for connection to the main grid. As noted in [1], the United States Department of Energy (DoE) aims to create an electricity infrastructure capable of supplying 20% of electricity demand using wind energy as the generation resource by the year 2030. Expansion of the existing network is necessary as the potential wind sites are remotely located with respect to the load centers. Given the implications of USF and expected significant levels of wind energy penetration, the motivation to accommodate variable USF in large transmission networks is justified.

Lack of a generic framework to generate the USF information regardless of the size of transmission network served as a secondary motivation. The intended extension needs significant automation and decision-making with minimal human intervention. Decision-making for this framework is in the context of selection of closed trails or loops in the network; estimation of loop flows; determination of multicollinearity issues; identify potential outliers; and, application of correction techniques to improve the quality of estimates. Accurate estimates of loop flows are crucial since they may be used in an economic model to determine the financial contribution of GENCOs towards the USF scenario in the network [2]. An algorithm to detect loops in the network regardless of its size is proposed. Formulation of the incidence matrix of the linear estimator with the loop information and line layouts synthesized by using geographical information systems (GIS) technique is also presented. Improvement in the quality of loop flow estimates using analytic ridge regression (RR) and robust regression (using M-estimators) is explored. Thus, a self-contained framework to compute loop flows within a contract period or planning period, as required is established in this dissertation work.

1.2   Objective of the research

The objective of this research is accommodating variable USF using estimates of minor loop flows (variable) due to large wind energy penetration in bulk interconnections. Investigating the impacts of stochasticity of wind energy on the market level accommodation from WGENCO perspectives is an additional objective. Conceptualization and demonstrations of the proposed framework are done on standard IEEE test systems with relatively simpler configurations as compared to bulk interconnections. Demonstration of results using test systems is an established practice and often resorted as related datasets are not readily available. From an email conversation with an expert in the field [11], and a dataset archival source [12], it can be inferred

4

that standard test systems belong to actual transmission networks as a subset. The availability of standard IEEE test systems enables researchers to simulate, test, and contribute towards the development of research on power systems.

Other objectives for developing the framework also included minimal human intervention in the estimation of loop flows that accommodate variable USF. The estimation framework was designed to be robust to be applicable to any network regardless of the size and complexity. This encompasses tasks such as selecting loops, forming a unique incidence matrix, assimilating variable power flows due to WGENCO contracts, estimating loop flows, accounting for WGENCO forecasting errors, and determining the individual contribution of WGENCOs towards total USF scenario. An algorithmic approach for detecting loops in any network was a major requirement for the framework. For execution of the mentioned sub-objectives the information (e.g., admittance matrix, GIS coordinates) already existing in the standard databases of power system networks was used. A crucial aspect of systematically interpreting the impacts of the stochasticity of power outputs of WGENCO was the adoption of an accurate probabilistic load flow analysis. Given the computing resources available, Monte Carlo simulations were chosen to simulate probabilistic load flows. In summary, the research objectives are:

- To develop a framework for accommodating variable USF resulting from the stochasticity of the outputs of wind power plants

- To formulate new techniques and enhance the accuracy of existing methodologies in estimating minor loop flow*s*

- To use existing components of standard databases such as GIS coordinates to automate the processing of the framework

- To investigate and establish the impacts of forecasting errors associated with the power outputs of wind plants on the loop flow estimates

- To improve the quality of estimates of loop flows by using advanced linear regression analysis

- To demonstrate applications of the estimates of loop flows from the proposed framework potentially applicable in practice.

In addition to techniques from power systems analysis, power systems economics and markets, myriad concepts from probability theory, regression analysis, stochastic systems, estimation theory, graph theory, and combinatorial optimization were adopted in this work. Figure 1 and Figure 2 provide information on the task-wise execution of the project and the milestones, respectively.

1.3   Analyzing USF in transmission networks

USF analyses have investigated transmission networks based on either models (e.g., linear estimation [2]) or practical experiences (e.g., study of particular interconnection analyzing congestion patterns and power flow patterns [6], [13], [14]). Historically, USF have been modeled and accommodated in multiple ways based on factors such as power transmission distribution factors (PTDF), transaction participation factors (TPF), interface participation factors (IPF), and line outage distribution factor (LODF) [15], [16]. These factors are established by the study joint efforts of GAPP and North American Electric Reliability Corporations (NERC) Mutli-regional Modeling Working Group (MMWG) with an objective to provide a solution on issues related to parallel flows for system operators and utilities. The primary motivation of the studies is the congestion of transmission corridors in the Eastern

interconnection that led to frequent curtailment for the trades. Shortcomings of the above approach in accommodation of USF are the assumption of linearity of power flows and the lack of thermal limit considerations. First contingency incremental transfer capability (FCITC) was introduced to include the thermal limitations on the transmission capacity as a modification [16]. Nevertheless, a suitable technique to model USF was needed due to the inaccurate assumption of applicability of superposition theorem to power flows.

Computation of circulating real and reactive power flows (analogous to loop flows) along with a severity index based on graph theory has been proposed by Davis, Weber, and Johnson [17]. Their approach models power systems network as a directed graph with the reference direction of the power flow on transmission lines determining the directionality of the line. The search algorithm is based on the depth first search (DFS) and the Tarjan's algorithm to determine a strongly connected vertex. This outward edge, with the maximum value of power flow, is used for determining a cycle and the MW flow associated with it. The computed value of MW flow in the cycle is assumed to be representative of the network. The limitations of this study are: the focus on unidirectionality of lines determined by *a-priori* knowledge and in not investigating any other cycles in the network to determine congestions. However, [17] provides a unique perspective of executing a generic analysis of circulating power flows using graph theory techniques.
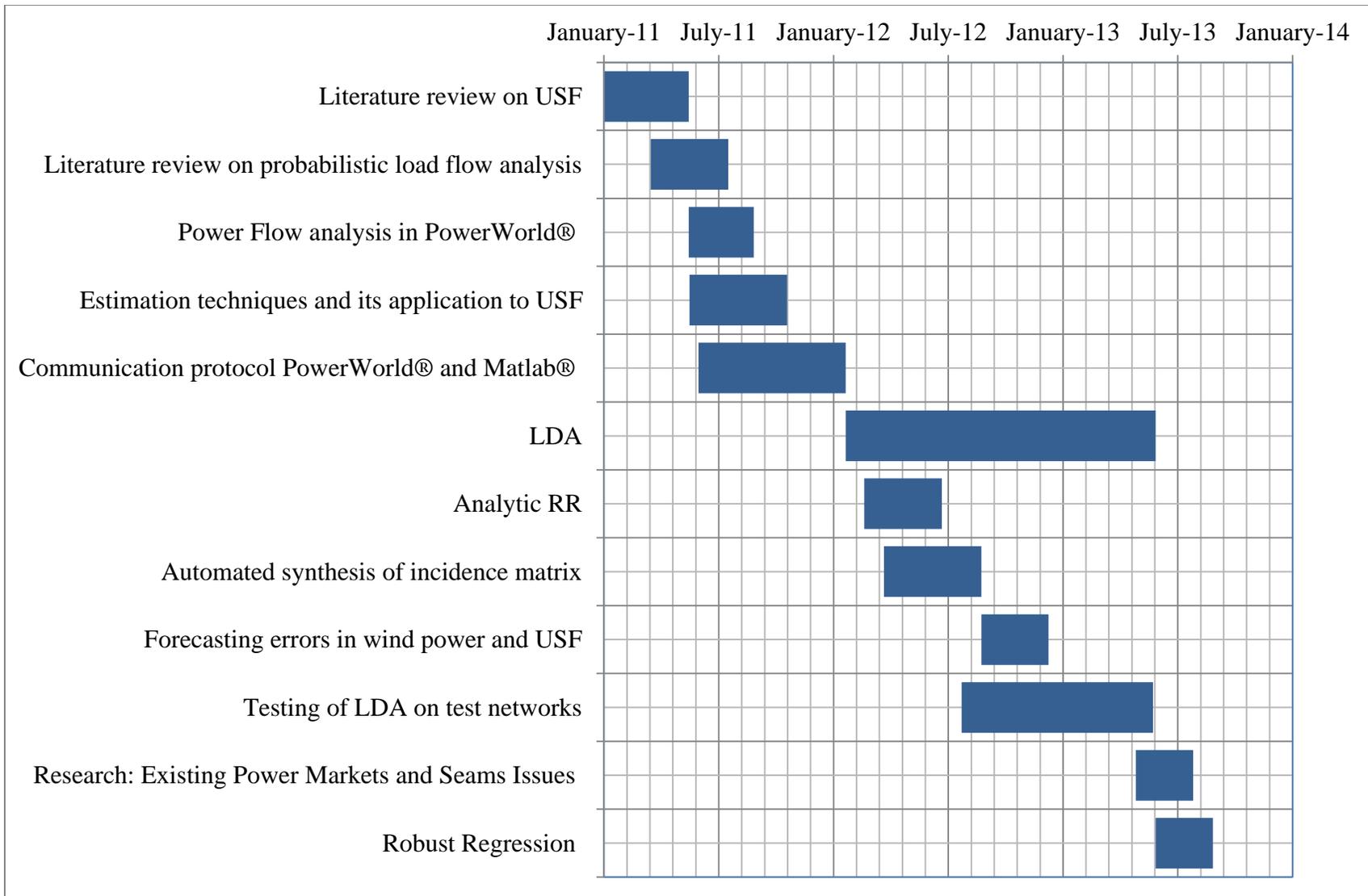
Figure 1 Timeline of the research

Figure 2 Timeline displaying the publications (top) and posters (bottom)

Legend:
1. FoPC: Frontiers of Power Conference
2. Greentech: Green Technologies
3. ASME: American Society of Mechanical Engineers
4. CREW: Center for Research and Education in Wind
5. PES GM: Power and Energy Society
6. T&D: Transmission and Distribution
7. NREL: National Renewable Energy Lab
8. LDA: Loop Detection Algorithm

10/17/2012
Use of GIS in the estimation of loop flows (FoPC) [35]

9/27/2013
Robust Regression (IEEE Trans. On Power Systems) [63]

1/14/2013
Impact of forecasting errors on USF (Greentech) [44]

6/11/2012
Analytical Ridge Regression Letter [56] (IEEE Trans. On Power Systems)

2/8/2012
Qualifying Examination

4/26/2013
Preliminary Examination

9/20/2013
LDA (IEEE Trans. On Power Systems) [34]

6/15/2011

10/14/2013

7/22/2011
Poster on framework of variable USF IEEE PES GM 2011

9/28/2012
Delivered a CREW talk [102]

8/5/2012
Poster on LDA with preliminary results

1/28/2013
Poster on impacts of forecasting errors on USF (NREL workshop)

5/8/2012
Poster on application of Monte Carlo Techniques to accommodate variable USF IEEE T&D 2012

11/14/2012
Poster on GIS applications in estimation of loop flows

9

Another graph theory based approach to interpret the impacts of phase shifting models on the values of loop flows was investigated by Cvijic, Cvetkovic, and Ilic [18]. The power systems network is represented using basic loops that are derived on the basis of the diakoptic algorithm built on the spanning tree formation of graphs. The analysis presented in [18] is based on dc power flows and hence obeys superposition theorem. Additionally, the transmission lines are only traversed as per the direction of the reference power flows. Applicability to networks of practical size is an advantage of the techniques presented in [18]. Determination of the contribution of generator company (GENCO) towards loads and line flows in networks with loop flows was analyzed by Das and Divan in [19]. Das and Divan use standard algorithms from graph theory to determine loop flows and determine flows at multiple paths using the *proportional sharing principle*. [19] also accounts for transmission losses occurring in the network for each GENCO. A linear estimator based modeling of USF using minor loop flows has been proposed by Suryanarayanan and Heydt in [20], [21]. Minor loop flows - a mathematical artifact for accommodating USF on the transmission lines is the crux of this approach. All the USF analysis presented so far are based on mathematical models fit to be applied to generic systems that lead to inferences regarding the impact studies.

The following set of non-model based studies summarizes the system specific impacts of USF observed in practice. Limiting of schedules, unavailability of paths, near stability operation of transmission lines as observed in the power systems of Pennsylvania – New Jersey – Maryland (PJM) interconnection, American Electric Power (AEP) System, and the members of WECC due to loop flows are reported in [22]. A unique case of the critical loop flow associated with the transmission lines around the Lake Erie is explained in [6]. Heavy congestion of transmission lines around Lake Erie was observed due to a change in the direction of loop flow

from the counter-clockwise to the clockwise direction. Enormous costs (estimated $96 Million) were incurred by all the market players to relieve the congestion caused by a specific transaction of energy by an unnamed entity in the NYISO was scheduled in the clockwise direction [6]. This indirectly led to higher prices of electricity with the reason being absence of a market mechanism to account for USF. According to FERC directives to counter this issue, NYISO investigated three potential solutions namely: buy-through congestion: congestion management: and, interregional transaction coordination. It was also pointed out by the then NYISO President and CEO Stephen G. Whitley that the solution of economic inconsistencies introduced by loop flows can be dealt within the deregulated market structure with suitable modifications [6].

Using qualified control devices (QCD) such as phase shifting transformers (PST) at crucial locations in networks to control USF, relieve congestions on transmission lines and exceptions to contributing schedules within the Western interconnection region is discussed in [23]. The location of qualified control devices and operational strategies is a classic multi-objective optimization problem. The objective functions for this optimization problem are – determining location of QCD, scheduling the operation of QCD, congestion relief, and regulating locational marginal prices (LMP). The principle and demonstration of the effectiveness of phase shifting transformers (PST) on a practical loop flow situation is provided by Lyman [24]. Fine and coarse resolution control over the phase angles corrections provided by PST are discussed with respective merits. Albaijat, Aflaki, and Mukherjee summarize a congestion study of the WECC system (for 2009 and 2018 cases) along with possible bottlenecks in transmission, hourly LMP, and the detection of possible congestion patterns in [13]. This study demonstrates a superior detection of congested transmission lines by using security constrained unit commitment (SCUC) instead of the optimal power flow (OPF). Thakurta, Hertem, and Belmans

11

discuss a system specific study regarding congestion on the European transmission corridors on account of significant penetration of wind energy in [14]. The objective function for this study is schedule PST for a day-ahead market scenario to relieve congestion in the Belgian grid with the constraint of minimum number of switching operations of the PST. An extensive bibliographical survey of technical work related to wheeling of power is provided in [25]. Studies related to specific wide area interconnections discussed so far indicate the existence and necessity of mitigating transmission line congestions under changing system characteristics.

1.4   Scope of the problem

The scope of the problem is described in this subsection to demarcate the proposed solutions and its applicability. USF were historically present in electric power systems; however, they became significant in the electricity markets only after the deregulation of the electricity industry. As mentioned earlier, USF can be defined as the unexpected flow of power on transmission lines from the markets perspective [2]. It can also be attributed to as a resource and infrastructure-scheduling problem. A few relevant concepts of power markets, economics and power systems operations are revisited to discuss the scope.

Primary energy markets operate on the basic principle of bid-sell, similar to any other commodities market. The bid and sell values of all the market participants provide the aggregated producer curves, consumer curves, and hence the market clearing price (MCP) [3]. The medium of trade in such markets is e-tags which are of a generic format [26]:

$$Point\ of\ receipt\ (POR) - Path1 - Path2 - Point\ of\ Delivery\ (POD), MW \qquad (1)$$

where, POR and POD refer to the physical nodes in the network with generators and loads respectively, $path$ refers to the set of transmission lines for completing the potential trade, and

MW indicates the committed value. The rights to access transmission lines — financial transmission right (FTR)— are distributed via an auction mechanism as well, and may be coordinated by 'Scheduling Coordinators (SC)' [27]. The ISO permits any such tentative trades based on the reliable system performance and operating limits. Typically, system performance and operating limits along with contingencies are gauged by computerized simulations based on optimal power flow, security constrained unit commitment, and contingency analysis [3].

A simplified example is presented in Figure 3, with three buses, two generators and one load. The generators connected at bus 1 and bus 3 representing two different GENCOs [2]. The line impedances for 1-2, 2-3, and 3-1 are Z_1, Z_2, and Z_3 respectively. Assume that GENCO 1 and the load have a bilateral trade or an approved e-tag for 100 MW. The FTR for the line 1-2 (FTR_1) is acquired by GENCO 1 for the intended trade of 100 MW. However, in practice GENCO 1 utilizes only P_{12} capacity of the line 1-2, which can be computed using the current division method (physical laws) for parallel branches. P_{12} is proportional to the impedance of the parallel branch (3-1 and 2-3) and similarly, P_{13} and P_{23} are proportional to the impedance of the line 1-2. Hence, the trade involving GENCO 1 and its identified load leads to USF of P_{13} and P_{23} on lines 1-3 and 2-3 respectively. This unaccounted use of transmission capability leads to the aforementioned network operation and market complications.

The USF on a line is computed as the difference between the expected flow (by adding all the FTRs associated) and the actual flow when all the approved e-tags are approved and operational. It is noteworthy that all the trades initiated by GENCOs contribute towards the total USF in the system. An important assumption in the case discussed is that the FTRs exactly match the two physical nodes in the network, i.e., the POR and POD match the generation and load bus respectively.

$P12 =((z2+z3)/(z1+z2+z3))*100$

$FTR\_1 = 100\ MW$

$P13 =((z1)/(z1+z2+z3))*100$

$P32 =((z1)/(z1+z2+z3))*100$
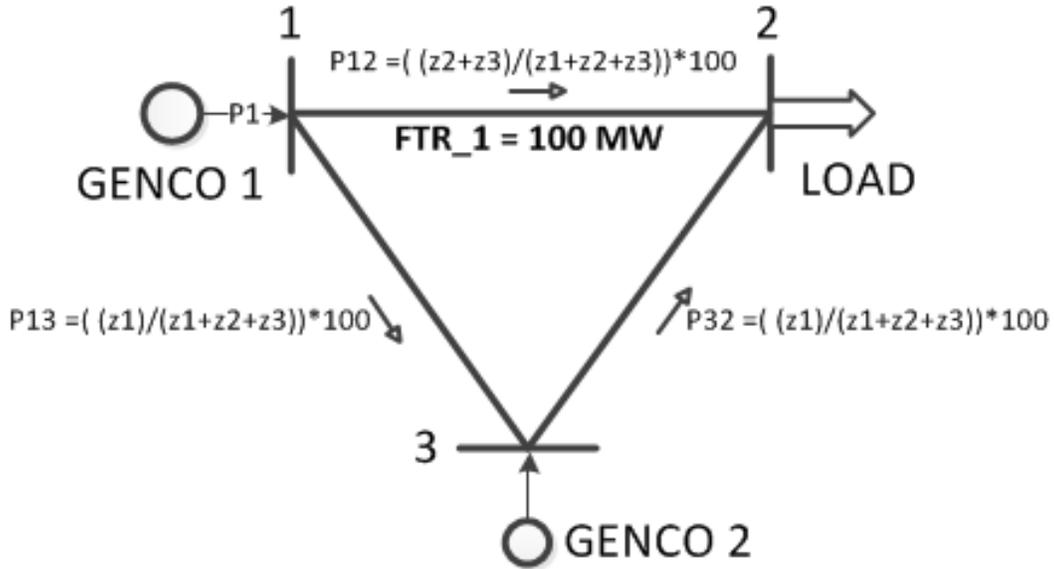
Figure 3: A simplified example demonstrating USF [2]

In practical wide-area networks this may not be necessarily true. The designation of POR/PODs depends on various factors such as available metering infrastructure, load zones, pricing schemes, market resolution, etc. In this dissertation, a one-to-one mapping between the set of POR/PODs and physical nodes of the network is assumed, thus making the proposed technique appropriate for the 'standard market design' (SMD), which is the objective of any completely deregulated power market [28]. This dissertation proposes techniques to accommodate USF for generators with variable output (wind turbines). Such generators influence the actual flow depending upon the stochasticity of wind profile and forecasting abilities. Thus a detailed study of accommodation of variable USF is undertaken in this project and documented in the dissertation.

1.5   Conceptualization of variable USF

The linear estimation framework to accommodate USF using minor loop flows as proposed by Suryanarayanan and Heydt, is adopted for the research in this dissertation [2], [29],

[30]. The prime reason for adopting the linear estimation is that the mathematical artifact of loop flows provides flexibility of application to USF accommodation techniques, and the accommodation obeys the non-applicability of superposition theorem to power flows. Equation (2) represents the linear estimator used to model loop flows:

$$[H]_{n \times p}[x]_{p \times 1} = [z]_{n \times 1} \tag{2}$$

where, H is the topology or incidence matrix, z is the vector of USF, x is vector of the mathematical artifacts - minor loop flows, n is the number of branches, and p is the number of loops [2].

1.5.1 Detection of loops in transmission networks

The linear estimation framework shown in equation (2) relies on visual decision-making. The selection of loops is based on visual interpretation of the topology of the networks, and visual selection of a bias value for ridge regression using traces. Visual decision-making cannot be applied to wide-area interconnections as these are topologically complex and lack precise line diagrams to choose accurate loops. The number of loops chosen to construct the incidence matrix should be a quintessential representation of all possible loops in the network. The number of loops that can be formed in a loop is prohibitively large if performed without limiting constraints. Additionally, all the graph theory approaches of determining loop flows discussed in the previous section [17]–[19] are based on treating the network as a directed graph with power flow serving as the reference direction. The loops detected using a particular reference direction holds true only for the particular set of generation and load conditions used to set obtain the reference direction. Thus, a significant portion of this dissertation deals with the development of a constrained algorithm that will synthesize loops in transmission networks using no *a priori* knowledge of power flows to assign directionality to lines. Additionally, changes in the topology

of a network occur on varying time scales. Networks experience changes in the form of addition of transmission lines that needs a few years of planning, whereas line outages can occur instantaneously. Instances such as transmission expansion or line outages will alter the USF scenario and the framework should recognize it along with related implications. The selection of loops in the network will be influenced by the changes in the network topology. The A-Star (A*) algorithm [31], the lowest ancestor theorem [32], and Dijkstra's algorithm [33] were used to develop the 'Loop Detection Algorithm' (LDA) that will synthesize closed loops in any network with a constraint of traversing an edge twice – once each in opposite directions. The admittance matrix of the transmission network (i.e., the connectivity matrix) serves as the input to LDA. The constraint of traversing any non-degenerate edge only twice bounds the search algorithm. The size and complexity of networks are the governing factors in the number of loops that can be formed - fewer loops can be formed in a network that is topologically simple and smaller in size and vice versa. Mohanpurkar, Zimmerle, and Suryanarayanan have summarized the LDA as a potential publication and it is under review at the time of writing this dissertation [34]. This paper documents the algorithm along with the order of complexity and a few examples of test networks with the detected loops.

1.5.2   Application of GIS techniques

The incidence matrix (H) in (2) maps the n lines to the p detected loops in the transmission network. The layout information of transmission lines can be assessed visually to allocate the elements of the incidence matrix; however, it is unsuitable for practical wide-area interconnections. The synthesis of the line layout information with respect to a preferred reference direction is obtained using the GIS coordinates of the buses in the network. GIS coordinates of the buses is a typical component of standard databases of power systems

networks. Assuming the transmission line as a straight line the *azimuth* is calculated with reference to north. A line not associated with a particular loop implies a zero entry in the incidence matrix at the respective location. For a particular loop, the magnitudes of *azimuth* of each line determine the element to be entered into the incidence matrix. An alternative technique of using the sequence of tracing vertices of the graph is also investigated. Both the GIS technique and the sequence of tracing vertices are completely automated and do not require human intervention. The incidence matrix and its accuracy of representing the mapping determine the accuracy of the loops detected. It is worth noting that the design of LDA and topology of networks impart a sparse nature to the incidence matrix. Mohanpurkar, Valdiviezo, and Suryanarayanan have published the role of GIS techniques in estimating loop flows in [35].

1.5.3 Wind energy in modern power systems

The earliest application of wind energy to perform mechanical work dates back 3000 years. The mechanical work performed by wind mills were primarily related to farming processes. In 1891, Dane Poul la Cour built the first ever wind electric conversion system (WECS) with a 18 kW power rating [36]. The evolution and implementation of WECS was sporadic and slow paced since its advent till the end of the 19th century. Although, the global installed capacity of WECS has grown exponentially from 1,480 MW in 1996 to 44,711 MW in 2012 [37]. A conservative estimate from 2012 to 2016 indicates a steady increase with global installations of WECS expected to reach the installed capacity mark of 59.24 GW [37]. Globally, efforts are underway to reach an optimum mix of generation profiles to reduce the harmful environmental impacts of conventional generation resources. As the manufacturing technology of WECS matured, the overall economics of wind energy has improved [38].

17

For the installation and commissioning of wind plants multiple criteria have to be considered such as, wind regime, technical feasibility, economic feasibility, transmission resources, availability of transportation services. The planning studies that determine the technical feasibility of projects has to account for the variability of the wind regime. Wind exhibits spatial and temporal variability [36]. Wind variation can be classified on a time range of a few seconds to a year. The temporal variability of wind has a greater impact on the performance and participation of WECS in the energy market. Correlation between the power output of WECS and load demand determines the load supplying capability of the wind power plant for a given contract period [39]. However, the power output is subjected to variability and uncertainty that cannot be controlled intrinsically, but only predicted.

1.5.4 Forecasting errors in wind energy

Unit-commitments by WGENCO are based on in-house forecasting models and are used to participate in the primary energy market [3], [40]. The outcome of forecasting models i.e., expected power output of WECS is intrinsically associated with errors. The accuracy of forecasting models depends on numerous factors – horizon of forecast, time resolution of the primary energy market, and the type of forecasting model. The horizon of forecast is suspected to cause the largest errors in the unit commitments of WGENCOs; the larger the horizon of forecast the larger are the errors [41].

It is a common technique to use standard distributions to model the forecasting error in power output of wind plants. Weibull, Beta [42], and the normal distribution [40], have all been employed to fit the forecasting errors in the power output of wind power plants. Forecasting errors for longer horizon span such as a few months to a year can be modeled using the normal distribution [40]. However, the distribution of errors associated with shorter spans of forecasting

18

such as a few hours to a few days demonstrates skewness and kurtosis, this requires modeling to use Beta distribution [41]. Prediction intervals are preferred over accurate estimates of power output of wind plants to reduce the risks associated with forecasting errors. The delta and bootstrap method of prediction in artificial neural networks are used to obtain prediction intervals with a known confidence level to demonstrate its effectiveness [43]. Thus, the investigation of forecasting errors in the power output of wind power plants on estimation of minor loops is crucial. Mohanpurkar and Suryanarayanan have discussed the impacts of prediction intervals associated with accurate point forecasts of output of wind power plants in [44]. This paper models the forecasting error in the output of wind power plants for an annual time period of simulation using the normal distribution.

1.5.5   Probabilistic load flow analysis

Power flow analysis is one of the most vital tools to extract the steady state information of the parameters in a network. Voltage magnitude (V_{bus}), angle (δ), real power (P), and reactive power (Q) are the four major parameters associated with each bus of a power systems network under steady state. Iterative and non-iterative techniques of performing deterministic power flow analysis have been documented in [45]. The deterministic power flow analysis assumes that the loads and generation values are constant in addition to no outages within network. A converged power flow solution provides the values of real and reactive power flows on the transmission lines of the network as well as the voltage profile on buses. The outcomes of the power flow analysis will significantly change in case of alterations in load levels or outages of network components. As noted in [46] by Sauer, a number of factors that impart an element of uncertainty to the parameters of importance (V_{bus}, δ, P, and Q) in power flow analysis are:

1. Line data uncertainty: Standard ambient temperatures are used to calculate the transmission line parameters e.g., a frequently used ambient temperature is 50° C. Thermal loading – a function of the actual ambient temperature – serves as a crucial limiting factor for the use of transmission lines. However, the choice of ambient temperature is a random process thus influencing the thermal loading limits of transmission lines in practice.

2. Load data uncertainty: Forecasts of load are always associated with uncertainty of the model itself. Additionally, the measurement of loads has an error which is a random process. The load demands have a high correlation with the temperature, thus attributing another cause of uncertainty to loads.

3. Generation data uncertainty: Forced outages of power plants are one of the prime causes of randomness on the generation side of power system networks. The generator data uncertainty as described by Sauer in [46] assumes that the input to the power plant is controlled, and hence the output is available as required.

The nature of output of generation techniques using renewable energy resources – wind power plants, solar panels, geothermal – are based on the nature of energy resources that are used as inputs. The energy resources harnessed in the aforementioned generation techniques are governed by random processes and hence have to be accounted as random variables. The framework of deterministic power flow analysis is not directly applicable with any of the uncertainties discussed. This necessitates the extension of existing deterministic power flow analysis to accommodate the uncertainty introduced by numerous factors involved in power systems networks.

Probabilistic load flow analysis was developed to account for variability of network parameters and random changes in network topology (or equipment outages) [47]. Probabilistic load flow analysis has found a wide range of applications in system planning and system operation in steady state. A detailed historical and current account of the applications of the probabilistic load flow is provided by Chun-Lien Su in [48]. The variables in probabilistic load flow analysis can be either extracted from heuristically or empirically created time-series datasets or distributions. As explained by Chun-Lien Su, probabilistic load flows can be simulated on either numerical or analytical bases. Numerical basis entails simulating all the probabilistic scenarios possible, and hence accounting the outcomes for the entire range of variability of random variables. The numerical approach is also known as the Monte Carlo simulation and is the most accurate method of simulation of probabilistic load flows as they retain the non-linear nature of power flow equations [48], [49]. The cost of computation is high and thus requires greater time for simulating probabilistic load flows using Monte Carlo simulations. Analytical methods deal with statistical techniques that use metrics of distributions of random variables involved (e.g., power output of wind power plant) and provide the outcome in the form of metrics of the resultant random variables (e.g., line flows) [48]. Analytical techniques to simulate probabilistic load flows using various bases such as the method of moments using the convolution theorem by Allan et al. [47], the Gram-Charlier type A series by Sauer [46], linearization of P, Q, V, and δ by Allan et al. [50], [51], fuzzy arithmetic methods by Alvarado et al. [52], and enhanced convolution theorem Schwippe et al. [53] are developed and verified on test networks. The analytical techniques need less computation time as compared to the numerical techniques of executing probabilistic power flow; however they provide less accurate results than the latter.

21

1.6   Solution of the linear estimator

The linear estimator shown in equation (2) represents an over-determined system of minor loop flows for accommodating USF. A number of solutions of iterative and non-iterative methods to solve the linear system is documented in [54]. One of the most commonly applied methods to solve equation (2) is based on the least squares minimization technique, known as the ordinary least squares (OLS) method. The incidence matrix (*H*) used to map selected transmission lines on the selected minor loops of the network may be ill-conditioned leading to numerical instability [55]. For instances of ill-conditioned incidence matrices, the OLS does not provide the most optimum unbiased estimates. Robust estimation techniques such as the ridge regression, robust regression, and Hubert estimation are recommended to counter the effects of multicollinearity [29], [30], [55]. The OLS method is characterized as an unbiased estimation technique [54]. In ridge regression (RR), selection of the bias value can be via visual methods (ridge traces) or analytical methods (auto-selection) depending on the applicability. An analytical RR method based on OLS estimates of loop flows is included in this research. Two advantages of this novel application are – that it counters multicollinearity of the incidence matrix, and it is suitable for a large number of regressors. The number of loops selected for wide-area interconnections is large; and this makes it inconvenient to plot ridge traces and choose a suitable value of bias factor. Mohanpurkar and Suryanarayanan proposed an advanced application of analytic ridge regression (RR) to counter multicollinearity issues in an incidence matrix for a test system in [56]. An automatic bias selection to improve the quality of estimates is featured in [56].

Building the regression model is an integral part of statistical experiments similar to the one presented in this dissertation. From the perspective of statistics, the accommodation of USF

22

using loop flows is a *controlled experiment*, as the analyst has the flexibility to select the regressors. Four major steps involved in building the regression model are – data collection, selection of regressors, model refinement, and model validation [57]. Data collection in the form of selection of loops, line layout synthesis, and power flow on lines is relatively easier. There are multiple feasible approaches to the selection of regressors that represents the underlying system optimally. Six criteria have been recommended to test, compare and select from possible alternative models representing the same experiment [58]. However, in cases of large number of regressors, there can be unyielding number of alternatives ($2^p - 1$, where p = number of regressors) to be tested. Performing the computation of a single or combination of the six recommended criteria for $2^p - 1$ different models may be tedious and time-consuming. Automatic search algorithms such the best subset selection, stepwise regression, forward selection, backward elimination exist to select the appropriate model with sufficient regressors. The use of a test-statistic (using t-distribution) to determine the statistical significance of regressors and hence selecting a subset may also be an acceptable technique [58]. Model reduction (MR) is used in this dissertation, based on the significance of t-stat of regressors. The process begins with the *full* incidence matrix with all the regressors. The procedure can be referred to as a stepwise reduction based on statistical and power system constraints as explained in chapter 4. The *reduced* incidence matrix is used for estimating the loop flows by robust regression (M-estimators). M-estimators are used to counter the impacts of potential outliers on the regression plane [59]. Model validation is performed using the standard techniques in literature [57]. The application of regression analysis is described on the IEEE-14 bus test system on account of its simplicity.

1.7 Statistical inferences in regression analysis

The incidence matrix in equation (2) is the as noted previously is a combined representation of the system observations i.e., loops and the network parameter i.e., transmission lines. The columns of the incidence matrix represent the regressors and are critical for the estimation techniques and outcome. The mutual dependence between regressors determines the quality of estimates, model adequacy, and applicability of the estimated model [57], [60]. Ideally, the regressors should be orthogonal to each other providing the optimum estimates of the coefficients of linear model. A common observation for most experiments is the mutual dependence between regressors and hence adoption of measures to counter the effects of multicollinearity. Selection of loops by LDA as explained in the section 1.3 is constrained by tracing each edge twice, once each in opposite directions. In other words, each transmission line will occur in exactly two loops that may lead to a departure from statistical independence between respective regressors. Additionally, visual techniques of loop selection may lead to multicollinearity issues in incidence matrices as reported in [2]. Thus, the solution techniques applied to estimate loop flows should be able to account for multicollinearity incidence matrices.

A standard procedure to validate the use of linear model is a part of linear estimation frameworks. Validation of applicability of linear estimator to a dataset is performed by using residuals of the regressors. Two ways to confirm model adequacy - normal probability plots of residuals lying on straight line and no discernible patterns of a scatter plot of residuals [57].

Empirical analysis is adopted to extract statistics of distributions of the estimated loop flows [44]. Comparison of a sample test distribution to a reference distribution and infer the relation between them is performed using the Kolmogorov-Smirnov (KS) test [61]. KS test can be employed to detect whether the empirically formed test distribution is sampled from the test

distribution. KS test is applicable to the continuous distributions only. For a pairwise comparison of two datasets, the KS test calculates the distance between the two distributions represented by the datasets and determines equivalency. The chi-squared goodness of fit is used to determine whether an empirically obtained distribution is a normal distribution. The test statistic is calculated for a known degree of freedom is compared with the desired level of significance [62].

1.8 Original contributions

The research proposed in this dissertation is original and contributes to multiple technical realms such as power system operations, power markets, and graph theory applications to power systems — all these areas are germane to the emerging *Smart Grid*. Multiple research articles based on the proposed research are published or are under review at the time of writing this dissertation in relevant significant archival sources [34], [35], [44], [56], [63]. The following are the brief contributions of the dissertation research:

1. Investigation and validation of the change in the nature of USF from deterministic to variable, in the presence of variable generation resources.

2. Development of a novel LDA based on existing graph theory techniques to detect loops in any network within reasonable complexity and memory bounds.

3. Development of an automated synthesis of the incidence matrix using the GIS coordinates and output of LDA.

4. Auto-correction of the quality of loop flow estimates in case of multi-collinearity issues of incidence matrix by using analytic ridge regression.

5. Investigation of the impact of forecasting errors in wind power in the form of prediction intervals on loop flow estimates.

6. Reduction of the impact of potential outliers in observations by using robust regression applications.

7. Applications of the proposed research in market and network level accommodation are outlined.

1.9   Assumptions

The contributions described in the section 1.8 are based on a set of assumptions. These assumptions are enlisted below:

1. The cause of variable USF on transmission lines is attributed to both variability forecasting errors in wind power.

2. Variable USF are modeled using a mathematical artifact *minor loop flows* via a linear estimator and is adopted from existing literature [2], [20]. The estimation of loop flows to account variable USF for an annual scenario is considered throughout this dissertation as a systems-planning study.

3. The linear estimator has no intercept and hence the regression plane passes through origin. This stems from the fact that, if there are no loop flows then the scheduled flow and actual power flow on all lines will match perfectly, thsu leading to an absence of USF.

4. Measurements of wind power at the point of interconnection are treated as the accurate point estimates. The wind power measurements have a one minute time resolution.

5. Each line in the transmission network is treated as a bi-directional edge and traversed exactly twice to form loops by LDA. The exceptions are edges associated with vertices with one degree vertices and *leaf* structures (refer chapter 2 for details).

6. GIS coordinates of the IEEE–14 bus test system are synthesized on basis of engineering judgment and explained in details in chapter 3.

7. Market expectations of power flows are computed using either of the two information sets – a) base case (chapter 3) or b) e-tags (chapter 4 and 5). Base case is used to compute market expected flows for planning studies to assess USF scenarios. In many instances, e-tags may not be available due to data privacy issues; hence the base case may be used.

8. The base case for the IEEE-14 bus test system is derived from the average values of distribution of load demands and the wind power measurements on an annual time scale. The base case in this dissertation is a representative scenario of the variability of loads and wind power.

9. E-tags are created manually and added arithmetically (through superposition theorem) in order to compute market expectations of transmission line power flows. E-tag detailing the schedule of a wind power plant is based on the average value of its annual output.

10. Normal distribution is used to model the forecasting errors in wind power for day-ahead forecasts. From a systems planning perspective, the prediction intervals of the forecasting errors are computed using the variance of the entire annual wind power dataset. However, both the horizon of prediction and the forecasting error distributions may change depending on the objective of the analysis.

1.10 Software tools

Considering the sources of uncertainty in power systems described in section 1.4 and the available computing facilities, Monte Carlo based probabilistic load flow was chosen for the analysis provided in this dissertation. Iterations within the Monte Carlo simulation are performed using the Newton-Raphson method on a commercially available software platform named PowerWorld®. The power output of wind plants is changed using a Matlab® script that communicates with PowerWorld® using the Simulation Automation Server (SimAuto) [64]. An electronic appendix is provided in the form of a compact disc with all the source codes and instructions needed to run them.

1.11 Literature search

The literature survey performed in this dissertation cites published work in the fields of power systems analysis [12], [15], [19], [22], [24], [25], [36], [53], [65], [66], power systems economics and markets [6], [13], [18]–[20], [64], [65], [67]–[73], probability theory and stochastic systems [41], [43], [47], [48], [50], [51], [74], [75] , estimation theory [2], [42], [57], [60], [76], graph theory and combinatorial optimization [17], [31], [32], [77]–[81].

1.12 Organization of dissertation

This dissertation is organized on the basis of the chronological development of the proposed framework of accommodating variable USF and related publications. The references [35], [44], [56] based on this dissertation are published at different venues and the references [34], [63] are under review. Figure 4 depicts the proposed contributions in this dissertation.

Chapter 1 provides an introduction to the research problem with an insight into the proposed dissertation. Chapter 2 is based on the intricacies of the LDA and derives directly from the manuscript under review, which is under review at the time of writing this dissertation [34].

Chapter 3 derives directly from [35] and deals with the synthesis of the incidence matrix (H) using a GIS technique and loop flow estimates (x_{ols}) using ordinary least squares. Chapter 4 derives directly from [56] and provides a framework for the application of analytic RR for countering multicollinearity in the incidence matrix. Chapter 4 is extended to include an application of building the regression model and robust regression (M-estimators) directly based on [63], which is under review at the time of writing this dissertation. Chapter 5 analyzes of the impact of forecasting errors associated with power output of wind power plants and is directly derived from [44]. Chapter 6 provides concluding remarks and the scope of immediate future work.

$$[H]_{n \times p} [x]_{p \times 1} = [z]_{n \times 1}$$

Requires 'loops' and 'line layouts'
1. Loops synthesized by the proposed LDA.
2. Line layouts are synthesized using sequence of tracing vertices and GIS coordinates.
3. Incidence matrix is auto-generated using step 1, 2 and reference directions for comparing with line layout.
4. Multi-collinearity detection and reconfiguration of incidence matrix based on statistical significance of regressors

Solution of the linear estimator should provide an estimate of loop flows. Estimation methods used:
1. Ordinary Least Squares (OLS) [Simplest solution].
2. Analytic Ridge Regression (RR) [Counters multicollinearity].
3. Robust Regression (M-estimators) [Counters the impact of potential outliers].

Actual flows are variable as the power output of wind plants is stochastic. Thus vector z, inherits variability from actual flows.
1. Monte Carlo simulations are executed in order to account for flows arising due to wind variability.
2. Actual flows are also associated with forecasting errors. Forecasting errors in power output of wind plants are modeled using normal distribution.
3. Market expectations of line flows are obtained from the super-positioning of e-tags

Figure 4 Figure depicting the proposed contributions in accommodating variable USF

CHAPTER 2

DETECTION OF LOOPS IN POWER SYSTEMS NETWORKS

This chapter presents the loop detection algorithm (LDA) to detect closed trails or loops in any transmission network. The algorithm is built on commonly used graph theory techniques and a modification of few specialized algorithms in combinatorial optimization. The input required for this algorithm is the admittance matrix of the network that is used as the connectivity matrix. The constraint of allowing the traversal of a non-degenerate line only twice in opposite direction bounds the search algorithm. The proposed algorithm utilizes a two–step iterative process with a second–order polynomial complexity. The algorithm is demonstrated on the IEEE–14 and the IEEE–30 bus test systems. The programming of this algorithm has been done in Matlab® and is being tested on numerous test networks. Few modifications are expected in the near future on account of reviews, but the crux of the algorithm is the same. This chapter has been created from the authors publication [34] and is under review during writing the dissertation.

2.1 Introduction

Historically financial settlements for bulk power systems in vertically integrated power system markets were characterized by simpler financial transactions. With deregulation, multiple players were introduced at all levels of the network and the bulk power market. In this scenario, transmission lines are owned and operated by entities known as transmission companies (TRANSCOs). As per the FERC Orders 888 and 889, these resources are equitably available to market participants such as generating companies – GENCOs and distribution companies – DISCOs [82]. Transmission rights are considered a fungible asset that is allocated, owned, and traded in the bulk power market. In the United States, the regional electricity market is managed by non-profit independent system operators (ISOs) [83]. ISOs ensure the maximization of the social welfare of the market by balancing the interests of all market participants, typically through open-access markets providing equal access to all GENCOs and DISCOs, while maintaining the security of electricity supply. All electricity transactions must be supported by sufficient resource allocation of transmission lines to route electricity from generation sources to load centers. Financial transmission rights (FTRs) help allocate and appropriately charge transmission users depending upon the transaction schedules, and consequently hedge risks of price volatility [70].

From the perspective of market management, a transaction is typically detailed in an electronic tag (e-tag) that consists of source(s) – GENCOs – path(s) or line(s), sink(s) – DISCOs – and the amount of power transfer (MW). The GENCO is expected to supply the load via the prescribed path only and the market mechanism assumes the path is physically and electrically valid. In practice, however, multiple paths are available for the power flow. Without flow control devices at each bus, it is virtually impossible to confine the power flow to the prescribed path.

Installation of a qualified flow control device at each source or even using services of an existing one is a costly option [84]. Hence, some portion of any scheduled power flow may traverse transmission resources that are not innate to an e–tag. These digressions are known as unscheduled flows (USFs) and result from the inconsistency between the physical dynamics of the bulk power system and the financial principles of market [29]. USFs contribute to multiple complex scenarios in the network such as line or path congestion and unexpected use of lines. The financial implications – i.e., costs associated with line utilization – are typically identified after generation and delivery contracts have cleared [29]. Hence, the modeling of, and accounting for, USFs is an imperative for maintaining an equitable market. The significance of inclusion of USFs and its financial implications is described in [85].

Some contemporary methods covering market and network accommodations of USFs for transmission networks are proposed in [14], [29], [65], [86]. USFs cannot be measured physically and they change with market conditions dependent on the market resolution. Hence, accommodation of USFs requires a flexible mathematical framework to account for this deviation between scheduled and actual flows. Using a mathematical artifact called *minor loop* – which are assumed to circulate in closed paths or loops in a network–and representing the USFs as a linear combination of the minor loop is a technique of modeling and accommodating USFs [87]. However, computing the minor loop flow paths requires thorough knowledge of network topology – specifically the identification of a unique and consistent set of loops within the transmission topology.

Loops are synthesized using one of the two techniques: visual and algorithmic synthesis. Visual synthesis involves a study of the pictorial form of the one–line diagram of the network and synthesizing loops using engineering judgment [29]. This technique is feasible only for small

33

networks that can be easily visualized and interpreted. For bulk interconnections at the continental scale, this technique is error–prone and laborious. Algorithmic loop detection is more viable for realistic networks. Advantages of algorithmic approaches include: a) accuracy; b) accommodation of special constraints by weighting busses or lines; c) no requirement of a visually manageable one-line diagram; and, d) no visual decision making requirement. Proof of convergence of the algorithms involves loop detection within finite temporal and spatial bounds. An algorithmic approach using constraints and weighting was applied for determining a feasible path for short–circuiting and hence de–icing conductors for a practical system in China [81].

The approach in [29], working from the GENCO perspective, provides a framework to accommodate USFs using the minor loop flows. The method proposes visual synthesis of loops and it is used to model a system matrix for the linear estimator. Loop flow estimates are eventually used in a contribution factor index so as to allocate an economic value to the USFs accounted for each GENCO with either a pay–or–charge outcome. The algorithmic approach in [86] estimates loop flows in a nodal frame of reference. The power systems network is assumed to be a multi-planar, directed graph with unidirectional edges. The technique utilizes the matrix tree theorem, forming a spanning tree by either depth – first search or breadth – first search to synthesize loops. The method constrains all loops to have at least one generator, which reduces algorithmic complexity by reducing the number of possible loops to the number of generators in the system network. Previously cited works estimate loop flows in a deterministic scenario, i.e., with an assumed fixed load and generation schedules. Reference [14] presents an application to the practical issue of optimizing the switching strategies of qualified controlled devices to control power flow in the European interconnect. The optimization objective here is to relieve congestion on the corridors due to increased wind energy penetration. USF analysis on a per–line

loading basis for power flow cases is performed with different levels of wind energy penetration, demonstrating an optimization methodology to effectively relieve congestion. Reference [17] explores the possibility of finding loops by partitioning the main network into sub-networks and treating the network as a directed graph. The method uses a constraint derived from the maximum flow branch included in a loop. A maximum loop flow is assumed to be representative of the respective sub-graph. The rest of the sub-graph information is unexplored and hence unaccounted. Identifying the transmission line user or alternatively coloring the network by using a dc power flow approximation is suggested in [65]. Users so identified are then charged using suitable rates or tariffs for transmission line use; this also helps determine strategies for flow control devices.

From the discussion above, it is apparent that the current loop identification methods rely, to greater or lesser degrees, on one or more heuristics to either identify loops directly or to reduce the complexity of identifying loops in the network. However, in graph theory, methods exist to detect graph sub-structures without relying on heuristic approaches. Below is a short review of some applicable graph theory foundation germane to the work presented in this paper.

An in-depth comparison and analogies between the power systems network and graph theory terminologies is provided in [79] that supports and substantiates the unique nature of connectivity in practical interconnections, which contain a relatively small fraction of highly connected busses and numerous sparsely connected busses. Reference [88] provides a primer on basic graph theory and a few advanced concepts. Reference [32] proposes an efficient method of finding a common ancestor between two points and demonstrates effectiveness by finding cycles (or loops) of negative weights. Some additional graph theory concepts related to edge deletion,

complexity analysis, and ancestor synthesis and storage are crucial to the algorithm presented in this paper and are described in [17], [32], [77]–[80], [88], [89].

The approach presented here builds on graph theory to present an iterative two-step network reduction technique that identifies minor loops using only a graph representation of a power systems network. The proposed algorithm is unbiased in assigning branches and directions to loops. Section II discusses the loop detection algorithm, and estimates the computational complexity and storage requirements. Section III demonstrates the algorithm on the IEEE–14 and IEEE–30 bus test systems, and discusses the respective loop synthesis results. Finally, Section IV discusses current applications and potential future work.

2.2   Loop detection algorithm

Graph description

The transmission interconnect is a graph where each bus is represented as a vertex, and each transmission line is represented as a bi-directional edge. The graph, $G(V, u)$, has v vertices, and u_{ij} or $u(i, j)$ edges such that, $u(i, j)$ is the edge representing the connection $v_i \rightarrow v_j$; similarly, u_{ji} can be interpreted as the connection $v_j \rightarrow v_i$. A transmission line between the buses v_i and v_j, is equivalent to the pair (u_{ij}, u_{ji}). The weight assigned to a unidirectional edge is 0.5 and subsequently, the degree of a vertex is calculated as 0.5 times the number of unidirectional edges. The graph G without any changes is referred to as the 'original graph/network', whereas any modified graph is referred to as an 'active graph/network.' Power system networks are typically sparse networks, i.e., the actual number of edges in such graphs is lower relative to the total possible edges. Additionally, these graphs typically do not have partitions due to the various technical constraints of power systems network operation and planning.

Consistency of graph

The algorithm requires that the graph remains consistent at all times during the algorithm. Consistency is defined by three conditions: a) no vertex may have a non-integer degree at the completion of any step of the algorithm; b) all vertices must have an equal number of out-bound and in-bound unidirectional edges at the completion of any step of the algorithm; and, c) no vertex is connected with the rest of the network with a single bidirectional edge in any active network. The second condition has to be specifically checked as a vertex may have an integer degree but an unequal number of out-bound and in-bound unidirectional edges. Throughout the algorithm, the cost function (weighting) of all the edges is equal to one, implying that no edge is preferable to any other when multiple options are present. Edges may be selected either sequential or randomly without loss of generalization.

Loop detect or reduce

Each step within the algorithm reduces the size of the graph by either eliminating a vertex from the active graph or detecting and removing a loop in the graph. All reductions are defined such that the resulting active graph is also consistent. Therefore, the algorithm may be restarted at any intermediate configuration, which is required for the iterative solution of certain graphs, as described below.

Macroscopic overview of the algorithm

Following is an overview of the major steps proposed in the LDA and are explained with sufficient details in the following subsection:

- **Step 1 Initialize:** Form the set of vertices and edges from the admittance matrix.

37

- **Step 2 Eliminate:** Vertices of degree one are deleted from the graph, as these busses have only one connection to the network and therefore cannot participate in loop flows. At the completion of this step, the graph has a minimum degree equal to two. We term this graph the "active network 1."

- **Step 3 Solution of vertices of degree two:** A vertex of degree two with at least one bidirectional edge – i.e., $count(adj_u[v_i]) \leq 3$ – participates in exactly two unidirectional paths. Vertices with degree two and associated edges can therefore be temporarily removed from the graph and replaced with the two unidirectional edges while maintaining the consistency and generality of the graph. We refer to this operation as "solving" the vertex. For typical transmission networks, iteratively solving vertices of degree two reduces the network size significantly. Vertices of degree two, where $count(adj_u[v_i]) = 4$ (i.e., vertices with four uni-directional edges) lack a bidirectional edge and have two potential solutions. Both the solutions are equally correct and may be required for the iterative solution described below.

- **Step 4 Identify the vertex with lowest connectivity:** Successive weighting using the degree of the vertices determines the vertex with low connectivity. The connectivity index of a vertex is calculated by adding the degree of the vertex with the degree of its adjacent vertices. The vertex with the lowest connectivity index is the starting vertex, v_{start}.

- **Step 5 Utilize a modified A* algorithm to identify a loop including the v_{start}:** Identification of a loop containing the starting vertex uses a wave search technique based upon the A* (A-star) algorithm for determining the minimum cost path to form a loop

[31]. A wave front search is initiated in an arbitrary outward direction from the starting vertex. The loop search is concluded when the wave returns to the starting vertex without assigning any vertex to two wave front levels. After constructing the wave front, a loop with minimum vertices is chosen by scanning the waves [32]. While scanning, one vertex for loop formation is chosen from each wave level such that an accurate loop is identified. This loop may or may not have the least possible vertices. Solutions of the vertices with degree two from step 3 should be accounted while forming loops. Edges involved in the loop are removed from the graph. Steps 3 through 5 are iterated until either the minimum degree of the graph is two or a terminating condition is detected.

Detailed description of the algorithm

This sub-section describes the above steps in detail.

Step 1 Initialize:

> Input (Y_{bus} ← admittance matrix); Define graph G:
>
>   Define all vertices, $v_i, i = 1 \dots V$
>
>   Define all edges from admittance matrix, u_{ij}.

The graph is derived from the admittance matrix of the power system network where each row or column corresponds to one vertex and each non-zero element of the admittance matrix represents one unidirectional edge. By convention and without loss of generalization, rows represent the "*from bus*" while columns represent the "*to bus*". The admittance matrix is symmetric, guaranteeing that each pair of connected vertices will be connected by exactly two unidirectional edges.

39

Explanation: The graph is derived from the admittance matrix of the power system network, which can be exported from commercially available software packages. Each row/column of the admittance matrix corresponds to one vertex of the graph and each non-zero element of the admittance matrix represents one unidirectional edge. By convention and without loss of generalization, rows represent the "*from bus* while columns represent the "*to bus*" to define edges. The admittance matrix is symmetric, guaranteeing that each pair of connected vertices will be connected by exactly two unidirectional edges.

Step 2 Eliminate:

```
FOR: i = 1:V

        IF: 𝒟(v_i) == 1

                  Delete v_i from G.

                  Delete u_{i*} and u_{*i} from G.

   ENDIF

ENDFOR
```

Solution of vertex with degree one:

Power systems network often contain nodes that are connected to the network by a single transmission line, such as generation stations, load centers or future expansion points. These nodes correspond to vertices of unity degree in G, and since the pair of unidirectional edges connecting these vertices to the graph cannot be a part of any loop, they are removed from the graph. Figure 5 demonstrates deleting a vertex of degree one, v_1, from G. In practical terms, a generating station at v_1 participates in loops *as if it were located at v_2.*
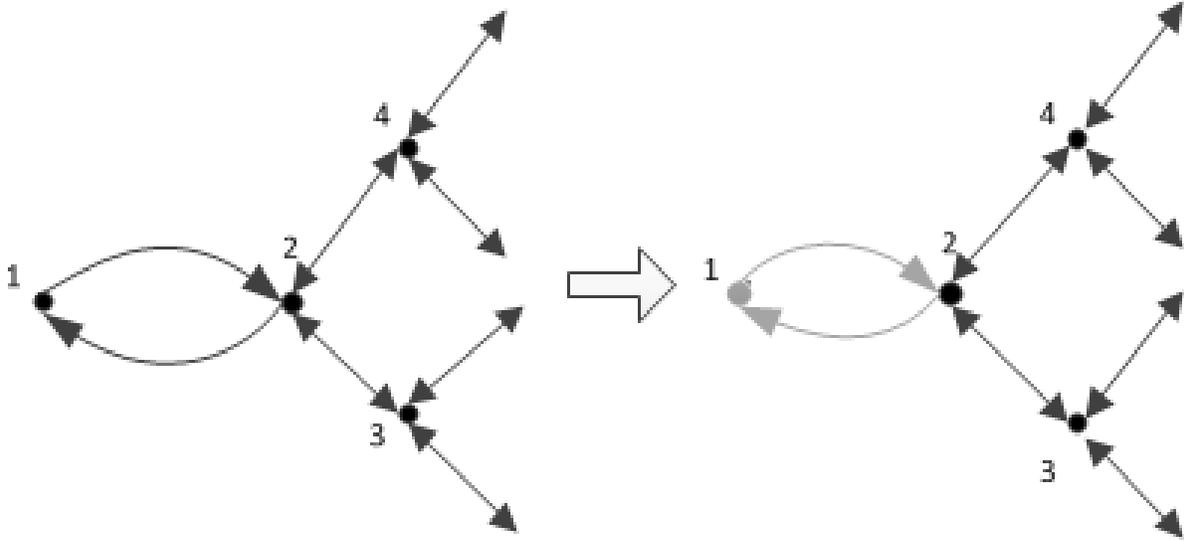
Figure 5 Deleting a vertex of degree one.

After Step 2, $\mathcal{D}(G) \geq 2$. For a vertex of degree two, there can be three edge combinations: two bidirectional edges, one bidirectional and two unidirectional edges, and four unidirectional edges. These combinations can be classified in two possible cases. First, in the case where the vertex meets the condition: $count(adj_u[v_i]) \leq 3$, it participates in exactly two loops. This participation can be recorded, removed from the graph, and later re-inserted when a loop is detected. Figure 6 illustrates the solution of a vertex with two bidirectional edges to neighboring vertices. Vertex v_3 is removed from G and two *composite edges* replace the connectivity through v_3: $\hat{u}_{134} = u_{13} + u_{34}$ and $\hat{u}_{431} = u_{43} + u_{31}$. No information is lost, the degree of vertices v_4 and v_1 remain unchanged, and the graph remains consistent, as per the definition given above. The general algorithm is:

41

```
WHILE min($\mathcal{D}(G)$) == 2 AND no terminating condition achieved

    FOR i = 1:V

        IF $\mathcal{D}(v_i)$ ==2 AND $count(adj_u[v_i]) \leq 3$

            Identify connectivity of $v_i$:

                $C_i = \{u_{ia}, u_{ib}, u_{ci}, u_{ci}\}$

            Form new edges, removing $v_i$ from G:

                $C_i' = \{\hat{u}_{cia}, \hat{u}_{cib}\}$

            Remove old edges and insert new:

                $G \leftarrow G - C_i + C_i'$

            Record $v_i, C_i$ for re-insertion in Step 5.

            IF a loop was identified

                Record loop

                Remove loop edges from $G$.

            ENDIFENDIF

    ENDFOR

ENDWHILE
```

If the graph previously contained another original or composite edge between vertices v_1 and v_4, at least one loop containing the new composite edge can be identified. For example, if edge $\hat{u}_{4*1}$ existed in the previous network state, a loop can be formed starting and ending at v_4: $L = \{v_4, v_3, v_1, v_*, \ldots, v_4\}$. The edges associated with the loop L are removed from G. Since the loop traces a path *through* all vertices in the loop, removing the loop from G removes exactly one

in-bound and one out-bound edge from each vertex in L. Therefore, G remains consistent during this operation.
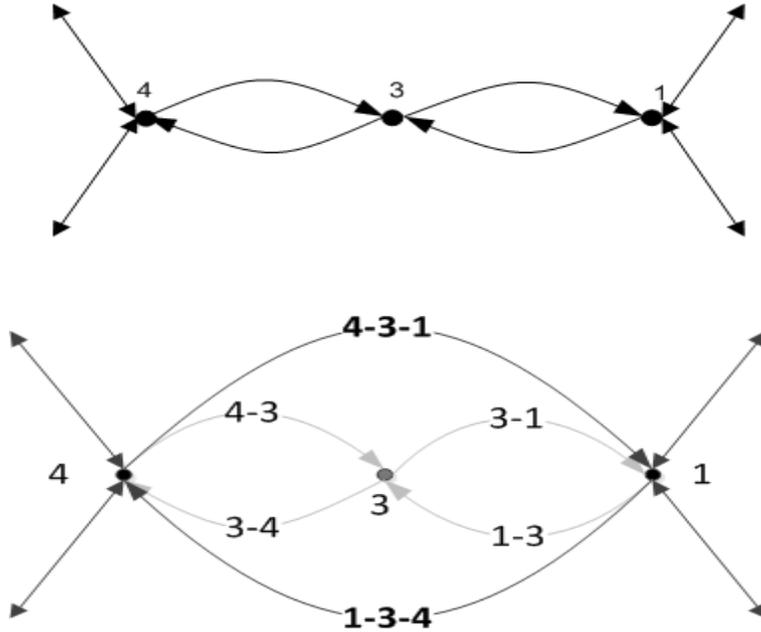
Figure 6 Solution of a vertex with two bidirectional edges.
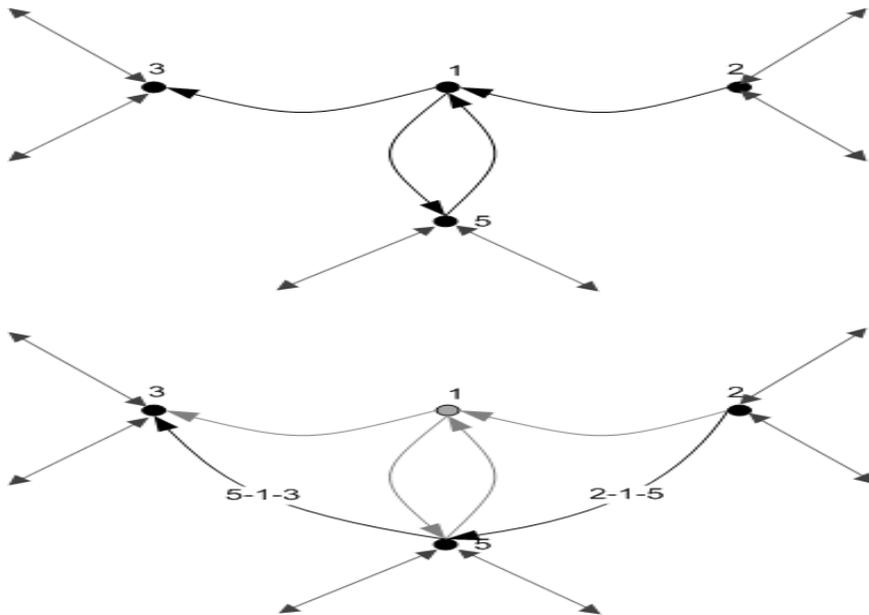
Figure 7 Solution of a vertex with one bidirectional and two unidirectional edges

Figure 8 illustrates the general case of four distinct unidirectional edges. There exist two possible solutions for the vertex v_1. Formation of the composite edges $\hat{u}_{413} = u_{41} + u_{13}$ and $\hat{u}_{215} = u_{21} + u_{15}$ or $\hat{u}_{415} = u_{41} + u_{15}$ and $\hat{u}_{213} = u_{21} + u_{13}$. The choice of either set of composite edges (a local decision) has an influence on the network reduction. Hence, the choice of composite edge and active network information must be stored till the network is reduced completely. If a later stage reduction produces a situation in which the graph cannot be reduced further, then the alternate local decision may be substituted for the solution, allowing the algorithm to proceed, producing an iterative solution of graph. These iterations are similar to those described in Step 5, below, where multiple loops can be detected during each step.
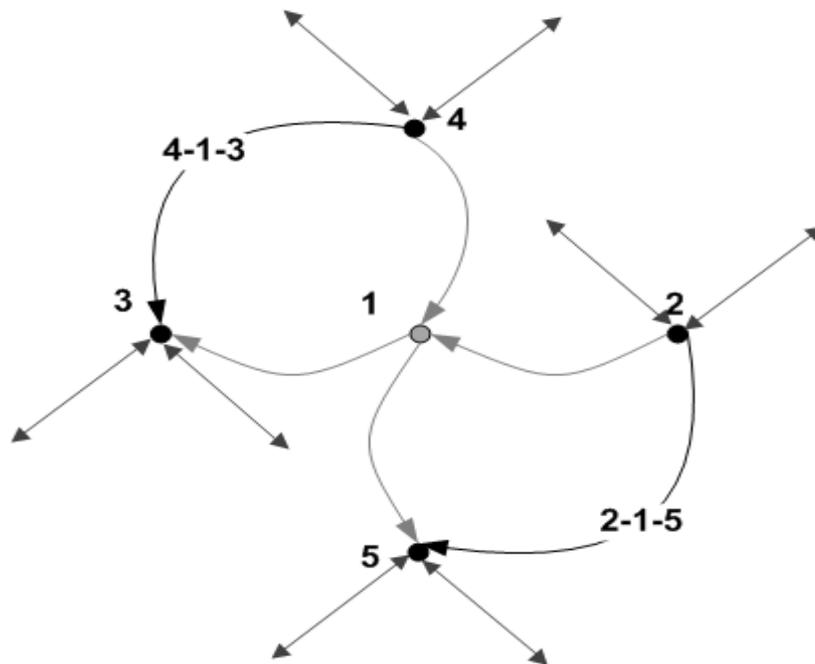
Figure 8 Solution of a vertex with four unidirectional edges

Steps 4 and 5 solve vertices with degrees greater than two by detecting and removing loops. Therefore, the algorithm is inherently iterative. We also refer the solution of vertices with degree greater than two as a "wave search."

Step 4 Identify the vertex with lowest connectivity:

A vertex with lowest local connectivity is detected using a standard technique of successive weighting. A starting vertex of the lowest local connectivity is chosen to reduce the number of wave-front levels required for loop detection. A connectivity index is calculated by adding the degree of a vertex and the degree of its adjacent vertices multiple (k) times. The vertex with a low index indicates a low connectivity to the remainder of the graph. Consider the graph shown in Figure 9, which has a minimum degree three. Figure 9 is also labeled with wave-front numbers, which will be utilized in subsequent discussion.

The connectivity index of vertex 1 is calculated by repetitively adding its own degree (3) to the degree of vertices 2, 3 and 6. Summation iterations (k) can be chosen heuristically. Choices between 5 to 15 work well in the example, and results are shown for $k = 12$. Vertex 6 has the lowest local connectivity [5371].
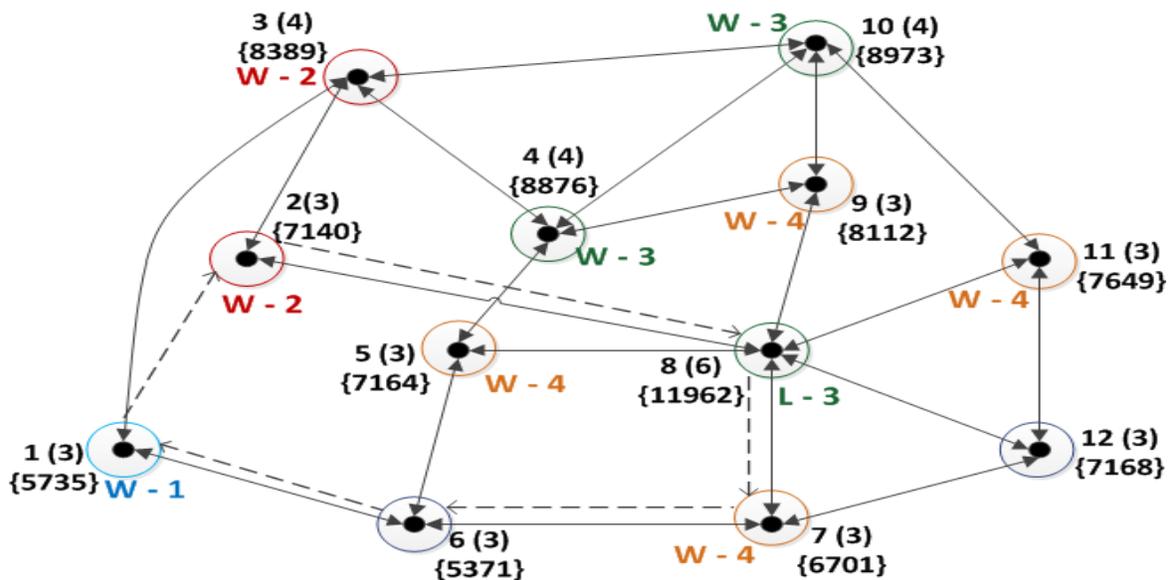
Figure 9 An example for determining the starting vertex and loop detection. Notation: 1 (3) → Vertex 1 (Degree 3); [5735]. → Summed weight = 5735. Dashed line indicates the loop detected by wave front.

45

Step 5 Construct a wave front: The algorithmic steps involved in constructing wave front to detect a loop are as follows:

Choose an outward direction from the starting vertex:

W_1 = {v_{out}};

Level of propagation: Level =1;

WHILE $v_{start} \notin W_{Level}$

      Level $\leftarrow$ Level+1;

      $W_{Level} \leftarrow$ adj$_u$[Traversed$_n$]

      Traversed$_{n+1} \leftarrow$ {W_{Level} - Traversed$_n$};

ENDWHILE

Store: Traversed, Level, W_{Level};

FOR I = 1:Level

      $loop$ = {$v_{start}, v_i, v_{start}$};

      (such that $v_i \in W_{Level}$ and check for composite edges.)

ENDFOR

The wave front, developed similar to the formation of a spanning tree, is initiated at the starting vertex by randomly selecting one out-bound edge from the starting vertex. The wave progression is guided systematically by the connectivity information through successive levels. Level 1 (W_1) is marked as the outward vertex; the adjacent vertices of the vertices from the previous level form the successive levels. A sequential list of vertices traversed by wave fronts is thus generated. The wave progression stops when it reaches the starting vertex with the loop being detected.

In Figure 9 with $v_{start} = v_6$, the edge u_{61} was chosen as the outward direction for constructing the wave front. Vertex 1 is labeled as the level 1 (W1). The next level of the wave front, level 2, contains vertices 2 and 3 i.e., $W_2 = \{v_2, v_3\}$ and level 3 is $W_3 = \{v_{10}, v_4, v_8\}$. This step of the algorithm continues similarly until, at level 5, vertex v_6 is again encountered. A modified lowest ancestor technique is used to form the loop with vertices from each wave level [32]. For Figure 9, the outcome of step 5 is the loop $\{v_6\text{-}v_1\text{-}v_2\text{-}v_8\text{-}v_7\text{-}v_6\}$. It is worthwhile to note, the total number of unique vertices in a loop so formed is equal to the total number of levels traversed. An alternative loop $\{v_6\text{-}v_1\text{-}v_3\text{-}v_4\text{-}v_5\text{-}v_6\}$ is an equally valid synthesis. However, if the wave front was initiated on the outward direction of vertex 6-7, then a smaller loop $\{v_6\text{-}v_7\text{-}v_8\text{-}v_5\}$ will be formed. Hence, the choice of the level 1 vertex influences the final outcome. It is common to observe multiple loops of minimum length formed during this step. All the possible loops are topologically valid, thus making the selection of valid loops arbitrary. The unidirectional edges used in the loop formation are eliminated from the graph, thus forming a new graph for subsequent steps. The algorithm iterates between steps 3 through 5 until either of the following two conditions is met: at least one vertex has a degree less than three; or the network is completely solved.

While any detected loop is a valid loop, there is no guarantee that the selection made at one step may not produce an active network configuration wherein a valid loop cannot be detected at a later step. If such a problem is discovered, the algorithm restores a previous state of the graph, selects a different solution path, and continues the reduction. No proof of convergence has been developed, but test cases indicate that valid solutions can be found with finite iterations.

Stopping criteria: The stopping criteria of the algorithm –i.e., exhaustion of loop information in the network – are checked by comparing the latest active network configuration with the possible remainders after the network reduction. The algorithm terminates when either of the two conditions occurs in the active network: a) all remaining vertices and edges are members of a single loop; and b) all remaining edges represent unidirectional links of some special cases of singly connected loops. If all the vertices in the active network obey the condition of $count(adj_u[v_i]) = 2$ and $\mathcal{D}(v_i) \leq 2$, then the first condition is met and the algorithm terminates. This network condition is referred to as the *inner link* and loop information is synthesized from it. The second condition represents topographical situations with no new information for loop flows, and can be disregarded. A set of vertices that may be defined as a *leaf* such that $count(adj_u[v_i]) = 2$ and $\mathcal{D}(v_i) \leq 2$ except for one vertex. However, the vertex with exception satisfies the condition $count(adj_u[v_i]) = 2$ and $\mathcal{D}(v_i) > 2$, and has to be identified in the active network no. 1. Graphically a *leaf* represents a single loop connected with the rest of the network at a single vertex and it contributes only one unique loop. The algorithm terminates when either of the conditions occur individually or in combination. Graphical explanations of these terminating conditions are included in section III. Figure 10 provides the complete flowchart of the proposed algorithm.
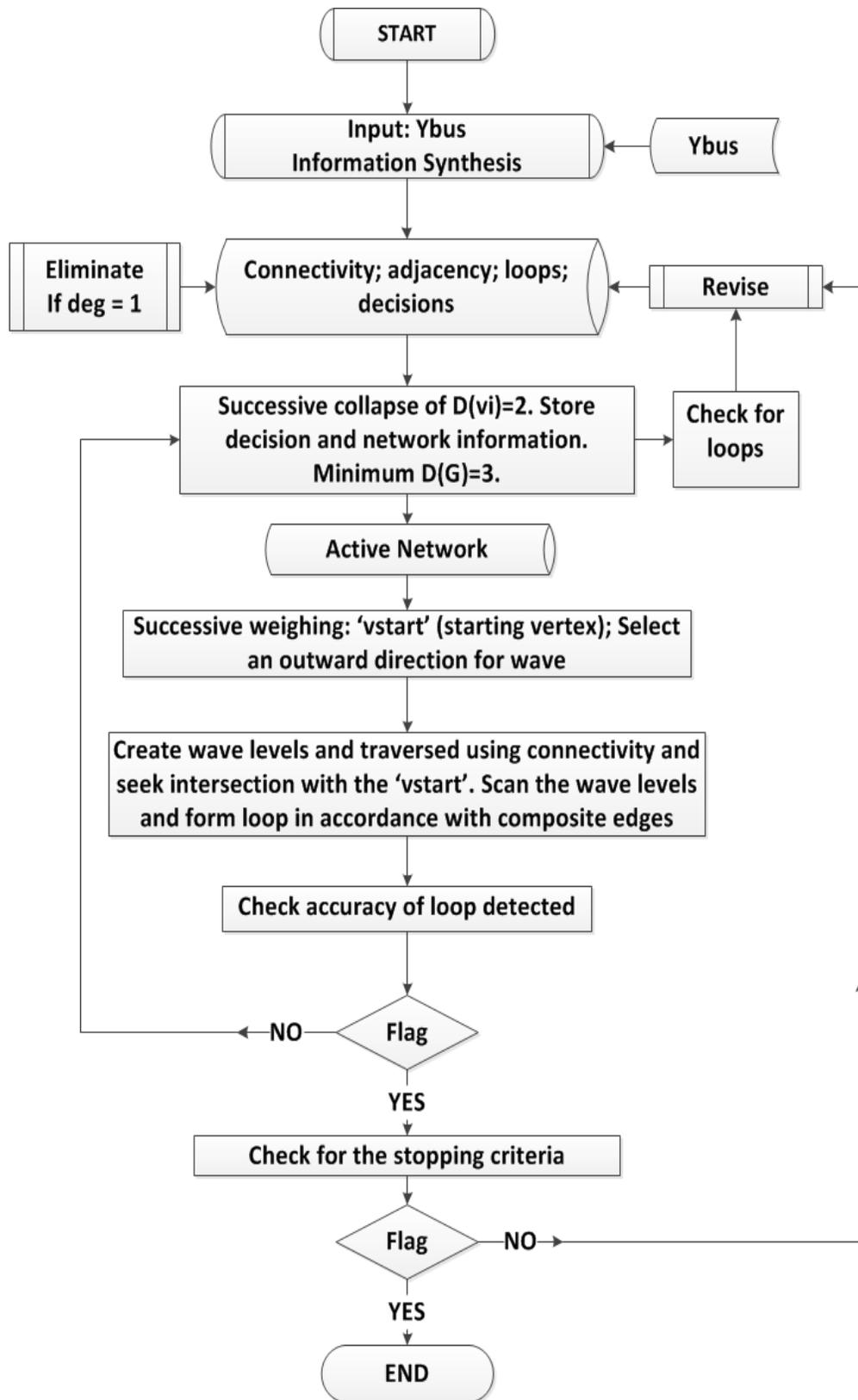
Figure 10 The flowchart of the loop detection algorithm.

Complexity Analysis

The computational complexity of the algorithm considers: 1) Solution of $v \; \forall \; \{\mathcal{D}(v \in G) = 2\}$; 2) Detection of a loop when $\mathcal{D}(v \in G) > 2$; and, 3) Traversing loops to re-insert collapsed vertices. A step-wise complexity analysis is undertaken starting with the step (1). In the worst case, the graph must be traversed entirely to find one vertex of degree two. Assuming minimal storage is required during the traversal, the computational complexity is therefore $\mathcal{O}_1 = \mathcal{O}(V)$. Solution of the vertex, when found, is a local operation, and thus the order of complexity of the solution is much less than the complexity of traversal. Step 2), above, involves two operations – assigning each vertex to a wave level, and the traversing back through the levels to identify the loop and remove it from the graph.

Assuming sufficient storage is available to track the wave level of each vertex, the complexity of assigning each vertex to a level is $\mathcal{O}_1 = \mathcal{O}(V)$. The number of wave levels is a function of the interconnections in the graph. The worst case is two binary trees traversed from each end toward a single point of interconnection as shown in Figure 11. The number of vertices at each wave level n is $V_n = 2^{n+1}$, and the number of wave levels, m, is a geometric series. Therefore, $m = \log_2(V + 2) - 1$. Considering the order of complexity, we get, $\mathcal{O}_2 = \mathcal{O}(\log_2(V + 2) - 1) \cong \log_2(V)$. To detect a loop, the algorithm must traverse through the wave levels by selecting a vertex at each wave level. If the maximum order of any vertex in the graph is l, then the number of levels and l bound the complexity of this operation. Noting that our theoretical binary-tree graph has two equal sides, we get, $\mathcal{O}(\text{loop detect}) = 2l \log_2(V)$. Finally, the order of complexity for step 2 is the sum of the traversal to number of the waves and the detection of the loop given by $\mathcal{O}_2 = \mathcal{O}(V + 2l \log_2(V))$. Since in practical transmission networks, $\log_2(V) \ll V$ and $l \ll V$, the bounding order of step two is governed by $\mathcal{O}_2 = \mathcal{O}(V)$.

The final step of re-inserting edges collapsed during the loop detection happens one time, and is bounded by traversing all unidirectional edges twice, or $\mathcal{O}_3 = 2\mathcal{O}(E)$. Since $E \leq lV$, $\mathcal{O}_3 = 2\mathcal{O}(lV) \cong 2\mathcal{O}(V)$.
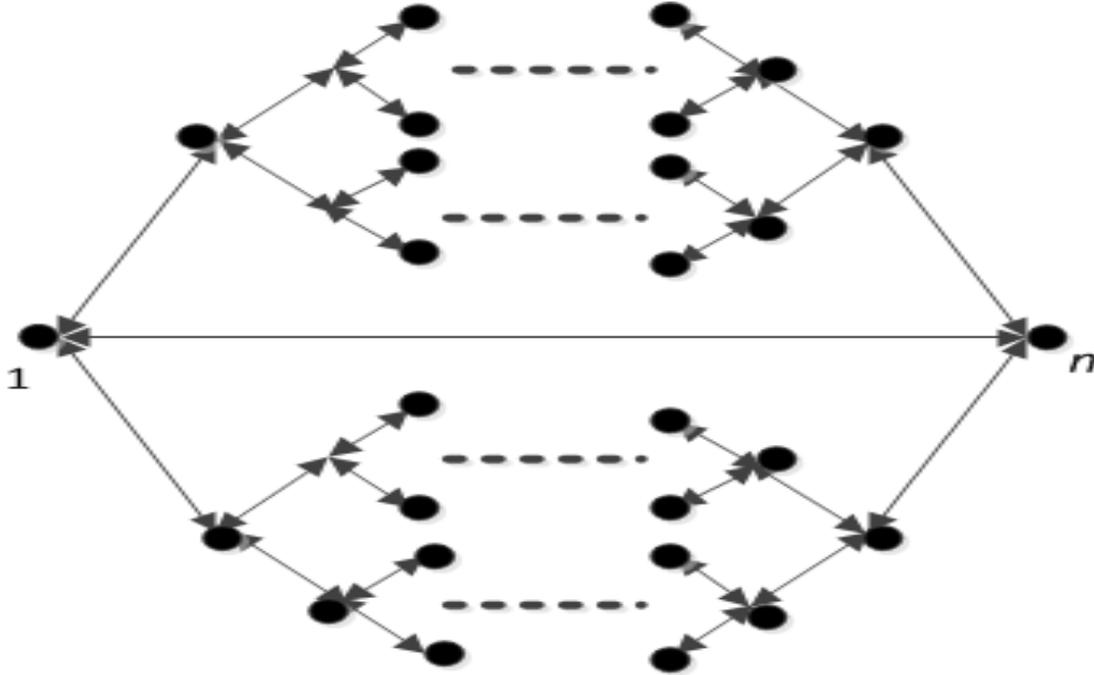
Figure 11: Binary tree with 2 distinct options for notional waves at each vertex (the dashed line indicates continuation of the binary tree as displayed).

Therefore, all constituent steps of the algorithm have an order of complexity of V. In the worst case, only one vertex of order two is reduced in each application of step 1), and only one loop is eliminated in each application of step 2). If we ignore that the loop detected will be on the order of m in length, each application of step 2) would at a minimum eliminate one path direction from $2m$ vertices, but will not eliminate any vertices. The two steps would then be executed, at most, N times, and the final loop annotation occurs once, producing an order of complexity for the entire algorithm of $\mathcal{O}_{total} < V(\mathcal{O}_1 + \mathcal{O}_2 + \mathcal{O}_3) < 2\mathcal{O}(V^2) + 2\mathcal{O}(V)$. Therefore, the algorithm is of 2nd degree polynomial order, i.e., $\mathcal{O}(V^2)$.

Since arbitrary choices are made in Steps 3 and 5, an iterative solution may be required. The total algorithmic complexity is then $\mathcal{O}(V^2) \cdot J$, where J is the number of iterations required. Estimating the size of J has not been completed, but due to the sparse connectivity of typical power systems we conjecture that J is bounded, and is of the same order of magnitude as V.

The required storage space of memory consists of: a) storage for the number of vertices in the graph, $S(V) + V * S(adj_b[v_i]) \cong 2S(V)$; b) storage to track the wave front, $S(V)$; c) storage for detected loops, or $2S(E) \leq 2S(lV)$, and d) storage of decision points and previous decisions for iterative solution in the form of bounded pointers, $S(2lV)$. Therefore the total storage is $S_{total} = 3S(V) + 2S(lV) + S(2lV)$, which is both bounded and reasonable for practical transmission networks.

2.3   Case Studies

The objective of the proposed LDA is to synthesize minor loops in transmission networks for use in multiple applications such as minor loop flow estimation, unscheduled flow accommodation, loss allocation, and variability analysis of renewable resources in the transmission network. The LDA is applied to case studies on the standard IEEE–14 and IEEE–30 bus test systems.

Case study on the IEEE–14 bus test system.

Information sets related to vertices and edges are extracted from the admittance matrix. According to Step 2 of the LDA, vertex 8 is removed to convert the original network into the active network 1 as shown in Figure 12. Figure 13 - Figure 15 demonstrate the stepwise reduction of the test system along with the emerging information on the loops. Table 1 lists the loops detected in the IEEE–14 bus test system.

Table 1: Loops Detected in the IEEE-14 Bus Test System

Loops detected in the test system	
{2 1 5 2}	{4 5 6 11 10 9 4}
{2 3 4 2}	{4 7 9 14 13 12 6 5 1 2 4}
{6 12 13 6}	{4 9 7 4}
{4 3 2 5 4}	{9 10 11 6 13 14 9}

Here the complete loop information is extracted from the first step of reduction by successive collapsing of vertices with degree two, thus demonstrating the potential of a major reduction of the network by a computationally simple step. It is pertinent to note that the loops detected here are identical to a visual synthesis of loops [35]. The algorithm stops when all the vertices in active network when conditions $count(adj_u[v_i]) = 2$ and $\mathcal{D}(v_i) \leq 2$ are obeyed as shown in Figure 15. The stopping condition mentioned is true for all the vertices in the network and there is no remainder in the network.

Figure 12 Active network number1 for the IEEE–14 Bus Test System. (Italicized vertices are collapsed or eliminated).

Figure 13 Vertices 1, 3, 7, 10, 12, and 14 are collapsed in the active network 1 and multiple loops are detected. (Italicized vertices are collapsed or eliminated).
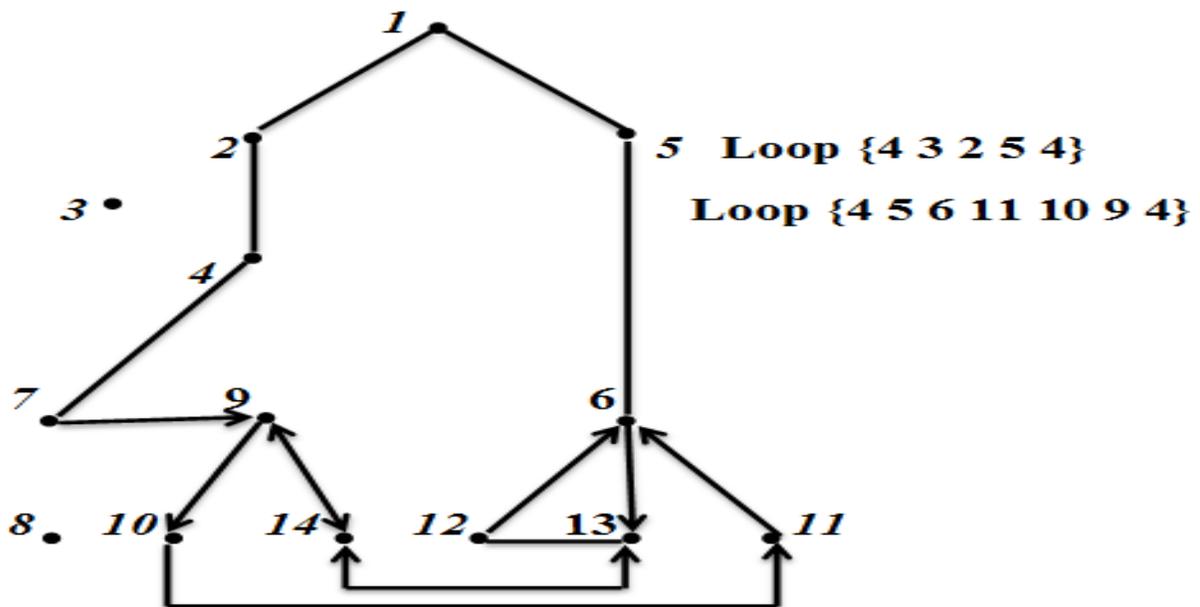
Figure 14 A subsequent active network showing loops identified by successive collapses of vertices with degree two. (Italicized vertices are collapsed or eliminated).
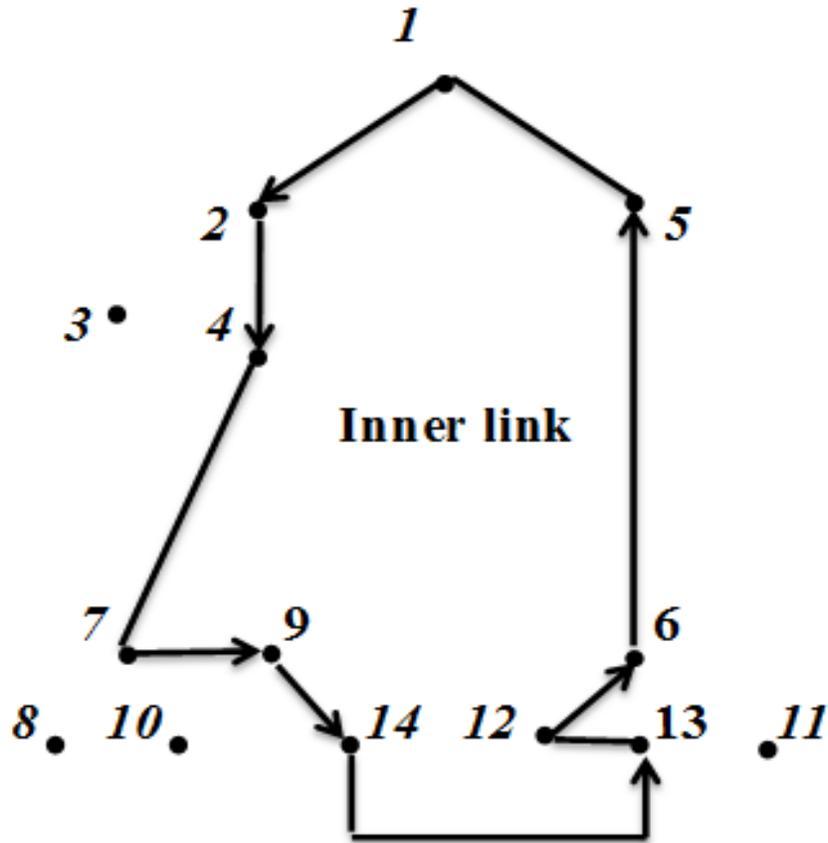
Figure 15 Final active network with the inner loop of the test system. (Italicized vertices are collapsed or eliminated).

Case study on the IEEE–30 bus test system

The IEEE–30 bus test system is topographically more complex than the previous case [29]. This network is reduced completely with the following sequential execution: a) successive collapse of vertices with degree two; b) solution of vertex with degree greater than two for detecting one loop; and, c) a final successive collapse of vertices with degree two. Eight loops were formed during the first successive collapse of vertices with degree two. The wave search is performed with the starting vertex as 6, with the outward path 6–10 chosen sequentially. The loop formed in this case is {6 10 12 6} with multiple edges collapsed between the vertices; this is

revised to {6 10 17 16 12 4 6} by inserting the collapsed vertices. Another successive collapse of vertices is executed to obtain the following loop no. 10, 11, 12, and 13 as shown in Table 2. The network is completely reduced at this juncture with all the loop information stored. The active network no. 1 of the IEEE–30 bus test system is shown in Figure 16. Figure 17 shows the active network no. 2 after the first successive collapse of vertices with degree two. Starting vertex 6 is chosen after for wave search as it demonstrates the least local connectivity.

Consistency of the graph at every step is a necessary condition for validity and is ensured for both the case studies. For example in the IEEE–30 bus system, the vertex 24 has degree three i.e., three bidirectional edges connecting to vertices 6, 10 and 15. Hence, three incoming arrows from the collapsed vertex 25 are shown in Figure 17. Both the conditions described in the stopping criteria of section II are observed in this test system. A *leaf* can be identified in the active network no. 1 with the set of vertices 27, 29, and 30 and the condition is checked as explained in the previous section. For the vertices 29 and 30 in the active network no. 1, the existence of *leaf* is identified by $count(adj_u[v_i]) = 2$ and $\mathcal{D}(v_i) = 2$. The vertex 27, is the vertex in the *leaf* that does not follow the connectivity conditions mentioned. In such case the reverse direction with vertices 27–30–29 highlighted by dashed lines in Figure 17 will be a remainder and cannot be used within any other loop. In addition to the *leaf*, an inner link is the remainder with vertices 10 22 24 23 15 18 19 20 10. The vertices in the inner link obey the condition $count(adj_u[v_i]) = 2$ and $\mathcal{D}(v_i) \leq 2$. Thus, the combination of both the conditions described as the stopping criteria are satisfied and hence the algorithm terminates. It is crucial to note, the network reduction cannot produce any new isolated vertices during the iteration. In other words, no new vertices with degree one can be created in any active network by the algorithm.

56

Figure 16 Equivalent graph of the IEEE–30 bus test system [29]. (Italicized vertices are collapsed or eliminated).

Figure 17 Active network no. 2 after the first successive collapse of vertices with degree two and 6 as starting vertex for wave search. (Italicized vertices are collapsed or eliminated).

Table 2: Loops Detected in the IEEE–30 Bus Test System

Steps used in detecting loops	Loop no.	Loops detected in the test system
First successive collapse of vertices with degree two	1	{2 1 3 4 2}
	2	{2 5 7 6 2}
	3	{6 8 28 6}
	4	{12 14 15 12}
	5	{10 21 22 10}
	6	{4 3 2 1 6 4}
	7	{6 9 10 6}
	8	{27 29 30 27}
Solution of vertex with $\mathcal{D}(v)>2$	9	{6 10 17 16 12 4 6}
Second successive collapse of vertices with degree two	10	{10 9 6 28 27 25 24 22 21 10}
	11	{15 14 12 16 17 10 20 19 18 15}
	12	{24 25 27 28 8 6 7 5 2 4 12 15 23 24}
Inner link of the graph	13	{10 22 24 23 15 18 19 20 10}

2.4   Concluding remarks

The proposed LDA relies on graph theory concepts and is applicable to any power systems network that adheres to the class of graphs as described. The computational complexity of the algorithm is of the order V^2 and the storage complexity is manageable for ordinary computational devices. The constraint to limit formation of loops is provided by the number of allowable traverses for each edge. The LDA has two major iterative steps for reducing power

networks into minor loops: successive collapse of vertices with degree two, and the solution of vertices with degree greater than two.

CHAPTER 3

APPLICATION OF GIS TECHNIQUES TO SYNTHESIZE INCIDENCE MATRIX

This chapter presents an application of GIS coordinates to obtain the transmission line layouts with reference to the north. Combining the line layout and loops (chosen either visually or algorithmically) the incidence matrix for the linear estimator has been provided. Variable USF are accommodated using the minor loop flows simulated for a test network using the synthesized incidence matrix. OLS technique has been used to estimate loop flows and discussed. Power flow is executed using the PowerWorld® commercial software package and governed by a Matlab® script. Some concluding remarks and future scope are also presented in this article. The chapter has been directly created from the reference [35].

3.1 Introduction

Advantages of interconnected power systems include increased system security and reliability of supply, lowered operating costs, and increased available transfer capacity (ATC). However, interconnections also introduce challenges such as propagation of system events over wider areas and unscheduled flows (USFs) of electricity [29]. USFs represent the deviation of the actual power flowing on transmission lines from the market-scheduled flows. Contractual agreements are made based on the fair market assumptions to optimize cost and operate electric grid at a desirable frequency. Deviation of power flow may occur due to rerouting of power flows due to inadvertent changes in the topology and may lead to forced participation of utilities and other assets that may not directly be involved in particular trades. USFs are known to reduce ATC, increase transmission losses with operation at or near stability limits, and complicate cleared transmission pricing [29]. USFs are mitigated by curtailment of schedules and/or deploying qualified control devices in the system.

However, an alternative method for handling non-critical levels of USFs is by accommodating it in the market after the market has cleared. One of the methods of accommodating USFs is the use of linear estimation of minor loop flow*s* [29]. This is done by employing a simple linear regression model to estimate the minor loop flows using the topology of the system, and the difference of the actual branch flows and the expected load flows [29]. A disadvantage of the method described in [29] is that the topology matrix that yields the relationship between the USFs and the minor loop flow*s* is obtained visually. While this may be sufficient for smaller electric grids and for proposing the accommodation method as a proof-of-concept, it may not be amenable to power grids of practical size. This paper proposes an enhancement to the accommodation method proposed in [29] by utilizing geographical

information systems (GIS). Additionally, this paper also shows the applicability of the accommodation method to scenarios of high penetration of a stochastic generation source, i.e., wind energy, in wide-area power systems. This is considered nontrivial as the volatility in generation due to wind may manifest in USFs; and, in order to accommodate the USF properly, there is a need for understanding this effect.

This chapter is organized as follows: section 3.2 provides a brief note on GIS applications in power systems; section 3.3 illustrates the prior work on the estimation of minor loop flows using USF measurements; section 3.4 describes the application of the GIS technique for synthesizing the topology (incidence) matrix for the above mentioned estimation problem; section 3.5 applies this technique to a test system, and section 3.6 describes the nature of variability in estimates as a function of wind penetration. Section 3.7 concludes.

3.2   Geographical information systems in power systems

GIS, which deals with collection, management, and presentation of geographical data, is a widely used technique in infrastructure planning and management. Adapting GIS to power system network planning and operational analysis possesses the potential of multiple advantages. GIS applications that provide visual representation of physical systems play a crucial role in determining the status and operational control of field devices. An application of GIS in planning of MV distribution network of open loop configuration is discussed in [90]. This is a prime example of multi-objective optimization problem in distribution network planning and is solvable by various optimization methodologies. GIS have also found applications in operations and management of transmission and distribution networks due to the spatial distribution of infrastructure [13], [69], [91]–[93]. The spatial and temporal functions are combined to detect distribution network faults in [91]. Two classifiers, i.e., linear discriminant analysis and logistic

regression, are then trained for evaluating the obtained fault conditions to provide real time information about faults. Essential steps to transform spatio-temporal functions of the network and time series data management to form a device data management system (DDMS) are discussed in [92]. A feature of integrating the visualization of real-time and historical events in the power system enables a more efficient system; and, this employs GIS. Another optimization of operations cost for MV distribution network is proposed in [69], that functions by using smart meters serving as data inputs, and automated metering infrastructure (AMI) serving as the communications media, whereas the GIS providing the interaction platform. For transmission networks, an online monitoring system is proposed using similar principles such as spatial functions, GIS information, and communication networks to construct intelligent power grids in [93]. Management of congestion in bulk interconnections by mapping lines using GIS applications and congregating the various cost functions involved in the transmission of power is presented in [13]. Congestion patterns, hourly nodal pricing, planning of transmission lines, and efficient integration of renewable energy sources are also discussed on the same platform. An exclusive interaction of system operators with the physical infrastructure via interfaces is developed based on GIS for efficient management of network infrastructure [94]. Thus, several applications in the implementation of energy management systems (EMS) in power system networks using GIS exist.

3.3   USF accommodation by loop flows

As mentioned in section 3.1, a linear estimator is used to estimate minor loop flows for accommodating the USFs. The linear estimator is given by the equation (2) and it is solved using the ordinary least squares (OLS) technique and the pseudo-inverse of the non-square matrix, $H,$ to obtain the estimates of the loop flow, $\hat{x}_{ols}$ as shown in (3).

$$[\hat{x}_{ols}]_{px1} = [H^T H]_{pxp}^{-1} [H^T]_{pxn} [z]_{nx1} \tag{3}$$

The system matrix, *H*, may be considered as a combined representation of the closed minor loops of the network and the bidirectional transmission lines (branches) of the network. The following are the rules assumed in [29] while choosing the loops in order to maintain consistency in the selection process:

1. Each transmission line (branch) is assumed to be a bi-directional edge such that it can be traced in both directions.

2. Nodes with degree one, i.e., nodes with only one connection/branch, especially generation nodes or sparsely located load points, cannot be a part of any loops.

3. Loops are chosen such that all lines other than single connection lines are traversed twice in opposite directions.

4. Loops with lower number of nodes are preferred over those with larger number of nodes.

5. There is no fixed sequence of choosing a starting node to form a loop, and hence the resulting loop may vary from analyst to analyst.

6. Loops are chosen solely based on the visual interpretation of the network.

The selection of loops may also be influenced by the manner in which the one-line diagram is drawn, making the accuracy and clarity of the one-line diagram a crucial aspect in the process. However, visual inspection suffers from three basic drawbacks:

1) Incorrectness in detecting loops for multi-planar graphs;

2) Lack of accuracy checks; and,

64

3) Inapplicability for topographically complex connections, i.e., practical bulk interconnections. For test systems of small to medium sized networks this technique is useful to especially analyze planning and operational scenarios.

Consider the following simple test system consisting of 9 buses, shown in Figure 18 [95]. Visual inspection is sufficient enough to detect the minor loops in this network on account of its simplicity and small size, and is shown in Table 3. For the network shown in Figure 18, visual inspection of topography and rules described lead to following loops.
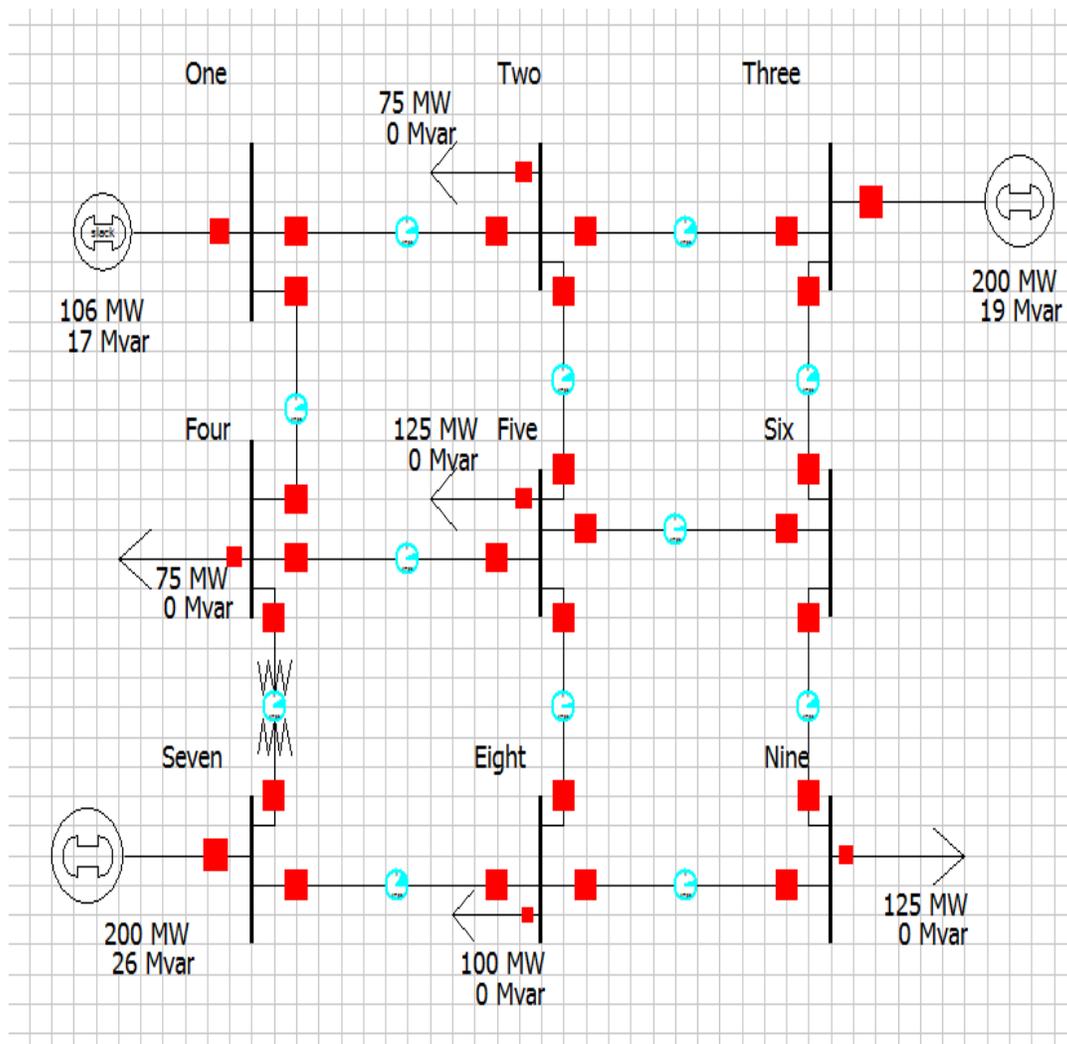
Figure 18 A notional 9-bus test system directly adopted from [95]

Table 3: Visual Synthesis of the Loops of the Notional 9 -Bus Test System

Loop number	Nodes of the system in the loop
Loop 1	[1 2 5 4 1]
Loop 2	[3 6 5 2 3]
Loop 3	[4 5 8 7 4]
Loop 4	[6 9 8 5 6]

Note that all the non-peripheral (observed visually) transmission lines (branches) are scanned exactly twice, going in opposite directions. For example the transmission line (branch) 2-5 is traced as 2→5 in loop no. 1 and as 5→2 in loop no. 2 as shown in Table 3. Using the loop information in the Table 3, the system matrix has to be synthesized; and, multiple methods can be employed to do so. In the first instance, the numerical inequality between the nodes of a branch can be explored. E.g., for the branch 2-5 in the loop 1 scanned as 2→5, the entry '+1' may be assigned whereas for the same branch scanned a 5→2, an entry '-1' may be assigned in the apt element of the topology matrix H. The logical operator approach uses the outcome of either '<' or '>' to obtain a pseudo-measure of allocating elements of system matrix. An alternative logic can be visual inferences of the location of transmission line layout with respect to reference directions. Visuals in Figure 19 clarify this element assignment convention.

This logic draws from the visual interpretation of the network layout and can be regarded as an extension of the visual synthesis of loops. The system matrix entries using the visual synthesis for the *Loop 1* (from Table 3) is given in Table 4. Table 3 provides the complete system matrix for the notional 9-bus test system using the discussed convention.

I.
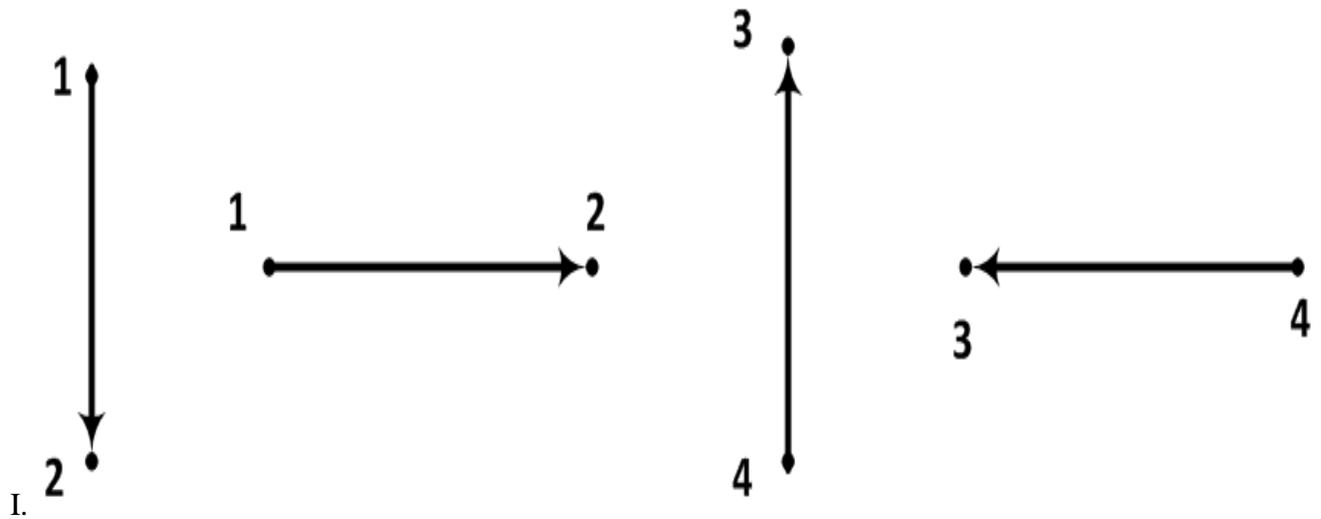
Figure 19 Scan directions along the reference directions 1-2 be allocated a '+1' and along the reference directions 3-4 '-1' will be allocated

Table 4: Visual Synthesis of the System Matrix Elements Corresponding to the *Loop 1* in Table 3

Branches	1-2	2-5	5-4	4-1
Elements	+1	+1	-1	-1

An enhancement over the above discussed procedure can be using each edge exactly twice in opposite directions as a criterion for loop selection. The peripheral edges 1-2, 2-3, 3-6, 6-9, 9-8, 8-7, 7-4, and 4-1 are scanned only once to form loops. Another loop can be formed by properly using all these unused edges. This loop represents the outermost periphery of the network and is not a minor loop. However, a topographically valid loop will be formed. The loop selection problem is limited to a finite solution by exhausting the direction of scan when used once.

Table 5: System Matrix for the Notional 9-bus Test System Using the Convention Shown in

Figure 18 [29]

	Loop 1	Loop 2	Loop 3	Loop 4
Branches				
1-2	1	0	0	0
1-4	-1	0	0	0
2-3	0	1	0	0
2-5	1	-1	0	0
3-6	0	1	0	0
4-5	-1	0	1	0
4-7	0	0	-1	0
5-6	0	-1	0	1
5-8	0	0	1	1
6-9	0	0	0	1
7-8	0	0	-1	0
8-9	0	0	0	-1

3.4   GIS application in system matrix synthesis

The above approach requires human decision-making and intervention in both the selection of appropriate loops and the consequent synthesis of the system matrix. The visual synthesis is not suitable for bulk interconnections due to complexity and large size; hence an automated mechanism is needed. In order to replace the latter with an automated process, a new technique of using GIS coordinates of the physical location of the buses (nodes) is proposed. For the GIS approach, the layout information of the lines is obtained by processing of the coordinates of the respective nodes for the branch under consideration. Other related information that can be synthesized using the GIS coordinates includes: elevation, distance, and angles with respect to a reference axis. Distances between buses can be assumed as approximately equal to the Euclidian line lengths. The layout information of the line is more significant in this regard to help synthesize the system matrix. A two-step procedure to automate the loop detection and system matrix synthesis suitable to bulk interconnections is proposed.

Step 1.   An algorithmic detection of loops in a power systems network.

Step 2.   Use the loops obtained from Step 1 to form the system matrix in an automated manner.

For the context of this paper, discussions pertaining only to the synthesis of the system matrix, i.e., Step 2, are given. Here, we assume that the input needed for executing Step 2, i.e., the nodes included in each minor loop, is known a priori. Discussion on the synthesis of the minor loops by algorithmic techniques is out of scope of this paper; however, such methods are currently under various stages of research and dissemination.

The dimension of the system matrix is (nxp), where n is the total number of branches and p is the total number of minor loops or regressors. Graphically, with respect to network, the rows represent the lines (branches) of the network in specific order, which may be obtained from a commercial software database used for creating the case information. The most common way of representing the line data is by using numerical values of the buses (nodes) connected in an ascending order. However, exceptions such as displaying or storing power flow only in the positive directions (as indicated by Figure 18) will lead to system-specific shifts in reordering branches. The columns correspond to the sequentially stored list of minor loops associated with the network (obtained from Step 1). The matrix elements corresponding only to the branches included in a particular loop will have a non-zero entry, whereas a branch not associated with a loop will have a 0 entry [1].

Convention #1: The visual synthesis of loops can also provide the directional layout of the line and hence can be used along with a proper convention to synthesize the system matrix. A convention for a line layout being *from the south to the north or from the west to the east* can be assumed to be a '*+1*', and vice versa as '*-1*'. This is a convenient method for small sized systems with accurate one-line diagrams and known directional information.

Convention #2: The intent of this paper is to replace the human decision-making of determining the line layout using the GIS coordinates. GIS coordinates are an inseparable component of standard bulk interconnection databases and are included from the planning stages. Using the GIS coordinates to obtain the *azimuth* for the line layout information is interpreted. The convention adopted here is based on the working database of respective networks (obtained either from standard datasets or commercial software packages) as: *from bus - to bus*. For this analysis the deviation of an actual transmission line from a straight line is ignored, i.e., the line

70

length of a transmission line is assumed to the Euclidean distance between the nodes. Transmission lines deviate from straight line due to multiple reasons such as available rights of way, topography of land, etc.

An *azimuth* is the angle, taken clockwise from north between any two points. The North Pole has an *azimuth* of 0º from every other point on the globe [94]. The interpretation of an *azimuth* is provided by an oversimplified example below: The GIS coordinates of point 1 are (latitude; longitude) = (21° 0' 0"; 89° 30' 0") and the point 2 = (40° 30' 0"; 105° 0' 0"). The *azimuth* of 319° is obtained for the straight line obtained by joining point 1 to point 2. The distance between the two points is 1632.342 miles. The distance so obtained might not be accurately equal to the line length as given in the database since deviation of actual lines from straight line is frequent. However, the intended purpose of setting up an incidence matrix does not require the distances between the buses at all, and hence the inconsistency between the geographical distance and line length can be ignored. The North Pole is used as the reference direction to obtain directional information. The loops obtained from the visual selection technique and the directional information are the two components used to synthesize H matrix Figure 20 below shows the convention used to allocate the elements of the system matrix.

In the loops chosen, for any branch of the network either of the nodes can be a reference point with the other point being the endpoint. The straight-line layout of this line segment with respect to the N-pole is obtained and the value of '+*1*' or '-*1*' is inserted at the appropriate element in the H matrix. A uniform convention has to be adopted if the line segment lies exactly on a boundary of the two regions i.e., coincides with the dashed lines shown in Figure 20.
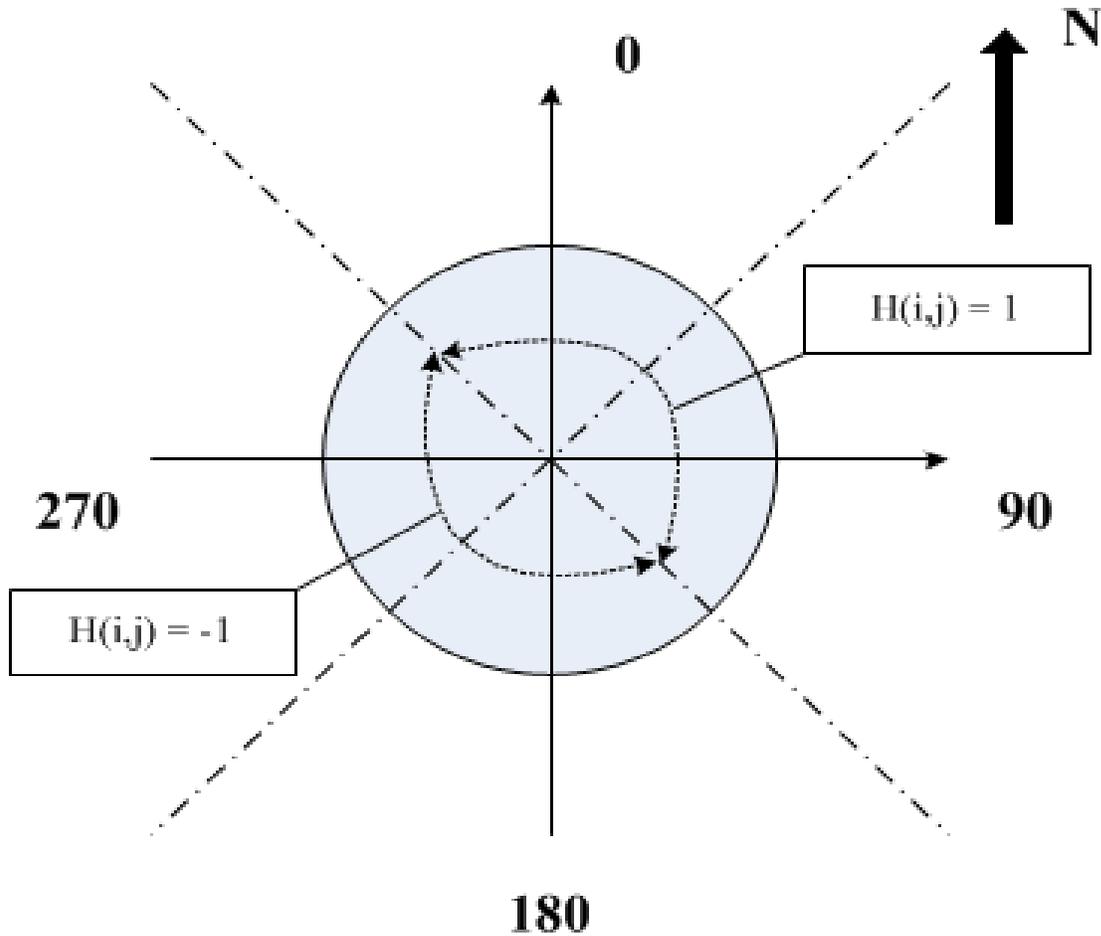
Figure 20 Convention for synthesizing the system matrix elements using *azimuth* (in degrees) of the edges in loops.

Table 6 shows that if the line segment has an angle of 135° then a '*+1*' will be allocated, which implies that the direction of the segment is westward which is positive. The exact reverse would be used for allocating a '*-1*' such that the segment layout is assumed to be southward. Similar argument can be made for a line segment with an angle of 315° will be allocated an element of '*-1*' assuming that the direction of this segment is eastward. Depending of the magnitude of this angle we are going to build the *H* matrix and use '*+1*' or '*-1*' according to Table 6, where x° is the *azimuth* of the line.

Table 6: Convention Adopted for Assigning System Incidence Matrix Elements Using GIS Coordinates and Directional information

Azimuth of an edge (x°)	Value
$0° < x° \leq 135°$	1
$135° < x° \leq 315°$	-1
$315° < x° \leq 360°$	1

All the loops of a network are processed such that each row of the H matrix will have only two non-zero entries i.e., '$+1$' and '-1'. Generation and load buses with degree 1 will have all-zero entries. Eventually these rows will be removed from the incidence matrix as they do not provide any notable information about loop flows.

3.5 Demonstration and discussions

The IEEE 14-bus test system is used to demonstrate the result of synthesizing the directional information of the branches in pre-selected loops. Using the angle (*azimuth*) of the transmission lines (branches) and the sequence of scan of the branches, a suitable incidence matrix is synthesized in Table 10 with the intermediate results displayed in Table 7 to Table 9. The pre-selected loops are obtained by visual inspection and using the rules of choice as explained in section 3.3. The test system comprises of 14 nodes, 20 branches, 6 generators, and 11 loads [51]. Appendix provides the details of the generator, load, and line flow values corresponding to the base case for the test system. The market expected line flows are directly assumed to be the base case line flows since they represent the most likely scenario. Pre-selected

73

loops are as shown in the Table 9. The standard database of the IEEE 14-bus test system does not specify the line lengths; hence, they are to be chosen rationally. Bus 1 is proposed as a point of reference to determine the latitudes and longitudes of the other buses (nodes) using assumed line lengths as a factor. In this example, the bus no. 1 of the test system is assumed to be located in Fort Collins, CO (40.60°N, 105.13°W). With these coordinates, and the assumed distances of each branch, we can calculate all the latitude and longitude for the buses (nodes). For this, we propose a *rhumb* direction for each one of the nodes, and with this direction and the distance we get the coordinates. A *rhumb* line crosses all the meridians of longitude at the same angle. The approximate line lengths assumed and the associated *azimuth* are shown in Table 7. Table 8 provides the network information along with the coordinate locations (derived from the assumed line lengths). The coordinates noted in the Table 8 are computed with the assumption that the system is located in the United States of America. Table 9 depicts the inclination i.e., *azimuth* (in degrees) of each edge in loops with respect to the North Pole measured in the clockwise direction, between the starting node and the ending node of the edge. These entries could have a value from 0° to 360°. The first row of Table 9 shows the *azimuth* values for the three edges in *Loop 1*, i.e., 2-1, 1-5, and 5-2. Sequential index is the total edges associated with respective loops. The largest is the 6th loop with 10 edges associated. Using the convention explained in Figure 20 and Table 9, the system matrix for the IEEE 14-bus test system is obtained as shown in Table 10. The system matrix has 38 non-zero elements. The total branches considered are 20; and, since the branch 7-8 (branch 14) is not traversed in any loop we will discard it. The remaining 19 branches have been traced twice in opposite directions, and hence we have the 38 non-zero elements. The system matrix has a sparsity index of 25%.

Table 7: Assumed Line Lengths in Miles and *Azimuth* Values

Node 1	Node 2	Line lengths (miles)	*Azimuth* (in degrees)
1	2	120	200
1	5	100	110
2	3	80	230
2	4	100	160
2	5	157	59.22
3	4	105	113.64
4	5	170	113.64
4	7	100	230
4	9	80	190
5	6	250	165
6	11	50	105
6	12	50	250
6	13	50	195
7	8	110	255
7	9	64	102.68
9	10	50	185
9	14	60	130
10	11	240	78.4
12	13	46	132.21
14	13	130	89.33

Table 8: Proposed coordinates of the IEEE 14-bus test system

Node	Characteristic components	Coordinates
1	Reference and slack generator	40.6º N, 105.13º W
2	Conventional Generator and load	38.96º N, 105.8º W
3	Load bus	38.21º N, 107.01º W
4	Load bus	37.59º N, 105.26º W
5	Wind farm and load bus	40.09º N, 103.35º W
6	Wind farm and load bus	36.58 ºN, 102.18º W
7	Wind farm	36.65º N, 106.64º W
8	--	36.22º N, 108.54º W
9	Load bus	36.44º N, 105.51º W
10	Load bus	35.75º N, 105.58º W
11	Load bus	36.38º N, 101.31º W
12	Load bus	36.32º N, 103.02º W
13	Load bus	35.87º N, 102.41º W
14	Load bus	35.87º N, 104.68º W

Table 9: Loops and rhumb values for edges associated with the loops listed sequentially in the IEEE 14-bus test system

Loop number	Nodes of the system in the loop	Sequential edge index									
		1	2	3	4	5	6	7	8	9	10
Loop 1	[2 1 5 2]	19.36	110.02	240.84	0	0	0	0	0	0	0
Loop 2	[2 3 4 2]	229.76	113.64	340.34	0	0	0	0	0	0	0
Loop 3	[6 12 13 6]	249.20	132.21	14.58	0	0	0	0	0	0	0
Loop 4	[4 3 2 5 4]	294.72	49.07	59.22	211.36	0	0	0	0	0	0
Loop 5	[4 5 6 11 10 9 4]	30.16	164.98	105.70	260.92	4.67	9.77	0	0	0	0
Loop 6	[4 7 9 14 13 12 6 5 1 2 4]	229.91	102.69	130.14	89.33	312.57	68.70	345.71	291.18	199.85	159.95
Loop 7	[4 9 7 4]	189.92	283.36	49.08	0	0	0	0	0	0	0
Loop 8	[9 10 11 6 13 14 9]	184.71	78.40	286.21	194.71	270.67	310.63	0	0	0	0

Table 10: System matrix for The IEEE 14 bus test system using the GIS coordinates

Branch number	Minor loop number								Branch number	Minor loop number							
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8
1	1	0	0	0	0	-1	0	0	11	0	0	0	0	1	0	0	-1
2	1	0	0	0	0	-1	0	0	12	0	0	-1	0	0	1	0	0
3	0	-1	0	1	0	0	0	0	13	0	0	1	0	0	0	0	-1
4	0	1	0	0	0	-1	0	0	14	0	0	0	0	0	0	0	0
5	-1	0	0	1	0	0	0	0	15	0	0	0	0	0	1	-1	0
6	0	1	0	-1	0	0	0	0	16	0	0	0	0	1	0	0	-1
7	0	0	0	-1	1	0	0	0	17	0	0	0	0	0	1	0	-1
8	0	0	0	0	0	-1	1	0	18	0	0	0	0	-1	0	0	1
9	0	0	0	0	1	0	-1	0	19	0	0	1	0	0	-1	0	0
10	0	0	0	0	-1	1	0	0	20	0	0	0	0	0	1	0	-1

3.6   Estimated USFs and variability

Experimental Setup

A Monte Carlo simulation is set up for the IEEE 14-bus test system in order to analyze the variability induced by the USFs as a result of penetration of wind energy sources. For this purpose, four wind farms of installed capacity approximately 222 MW, 99 MW, 95.9 MW, and 49.8 MW are assumed connected at buses 5 and 7 (one each) and at bus 6 (two numbers), respectively. The physical distances between the wind farms are approximately equal to the assumed line lengths. The wind farms are assumed to have local voltage regulation at respective buses. Real power injection by the wind farms measured at the point of interconnection[†] for a year is used as the variable input to the power flow. All other parameters pertinent to the power flow algorithm are maintained constant to execute the Monte Carlo simulation. Approximately 500000 iterations of power flow are executed to obtain the line flows as the output with wind farm output varying according to the obtained dataset. The base case is drawn using the average values of wind farm penetration and the fixed load and conventional generation values as tabulated in the appendix. The slack generator is expected to supply the transmission losses as well as serve as the secondary market for wind farms to supply deficits or surfeits, if any. Ordinary least squares (OLS) estimates are estimated using the Pseudo-inverse technique as shown previously.

Variability in USFs

Following are the histograms for the estimated minor loop listed in Table 9 under the annual heavy wind energy penetration scenario. Figure 21-Figure 24 display the histograms of some of the estimated loop flows, for *Loops 1, 5, 7, and 8*, respectively, as a representative

sample. The plots indicate a pattern of having a concentrated probability near specific MW values for the chosen inputs. The value of this USF corresponds to the most frequent penetration of the wind farm in the networks. Another observation about the wind farm outputs and the USFs is no fixed cause and effect relation. USFs and wind farm penetration exhibit both positive and negative correlation. Table 11 shows the values of correlation found between the different estimated minor loops and the wind farm penetrations obtained from the varying input. This is highly specific to the system and the inputs chosen; this result may vary depending on input choice or system choice. No generic results can be drawn from this analysis as yet.
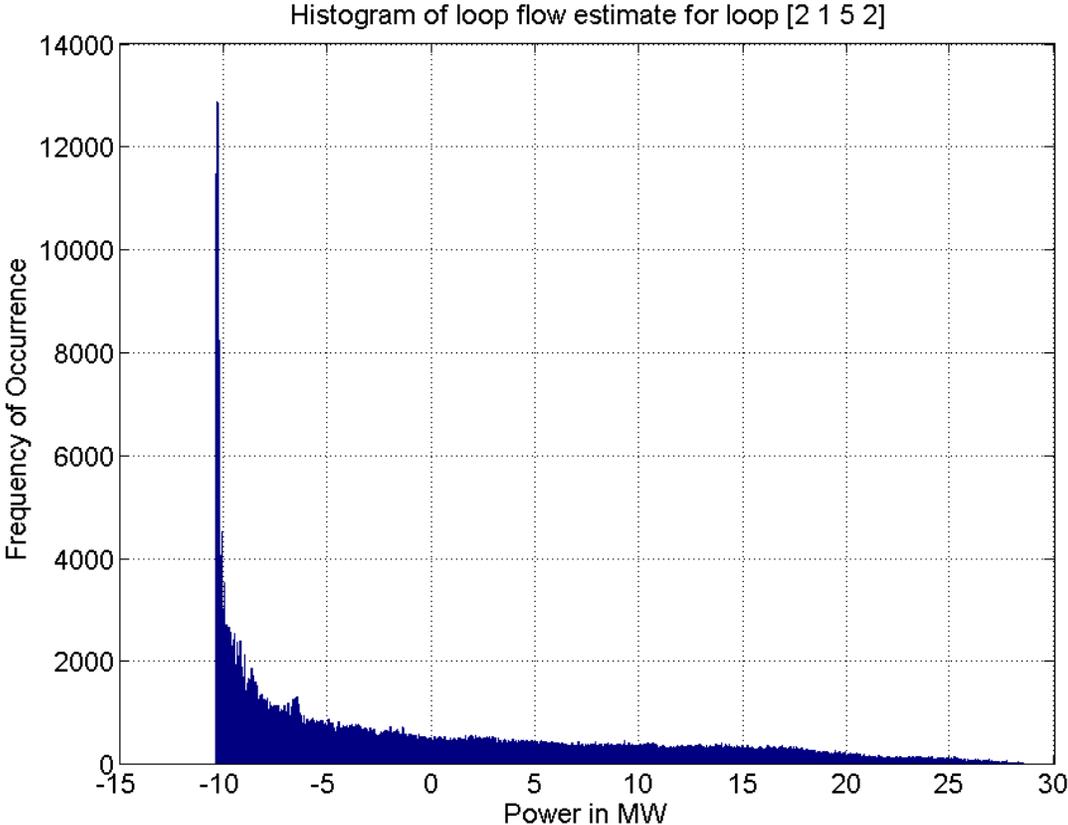
Figure 21 Histogram of estimated loop flow for the *Loop 1* using OLS.

Histogram of loop flow estimate for loop [4 5 6 11 10 9 4]
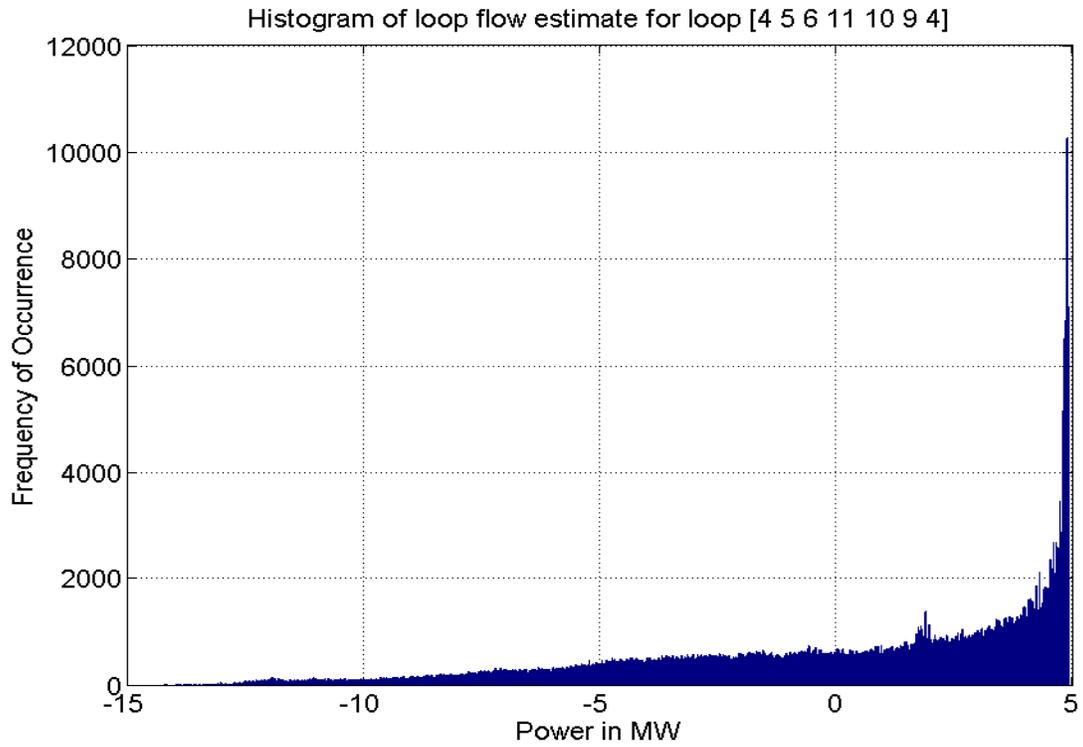
Figure 22 Histogram of estimated loop flow for the *Loop 5* using OLS.

Histogram of loop flow estimate for loop [4 9 7 4]
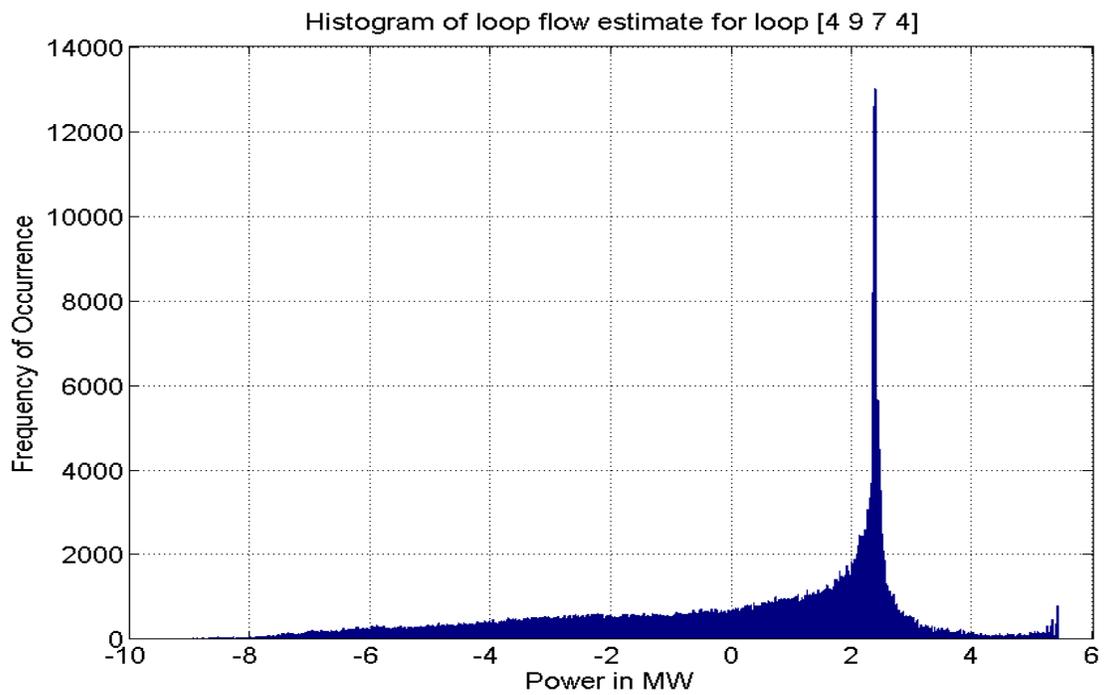
Figure 23 Histogram of estimated loop flow for the *Loop 7* using OLS.
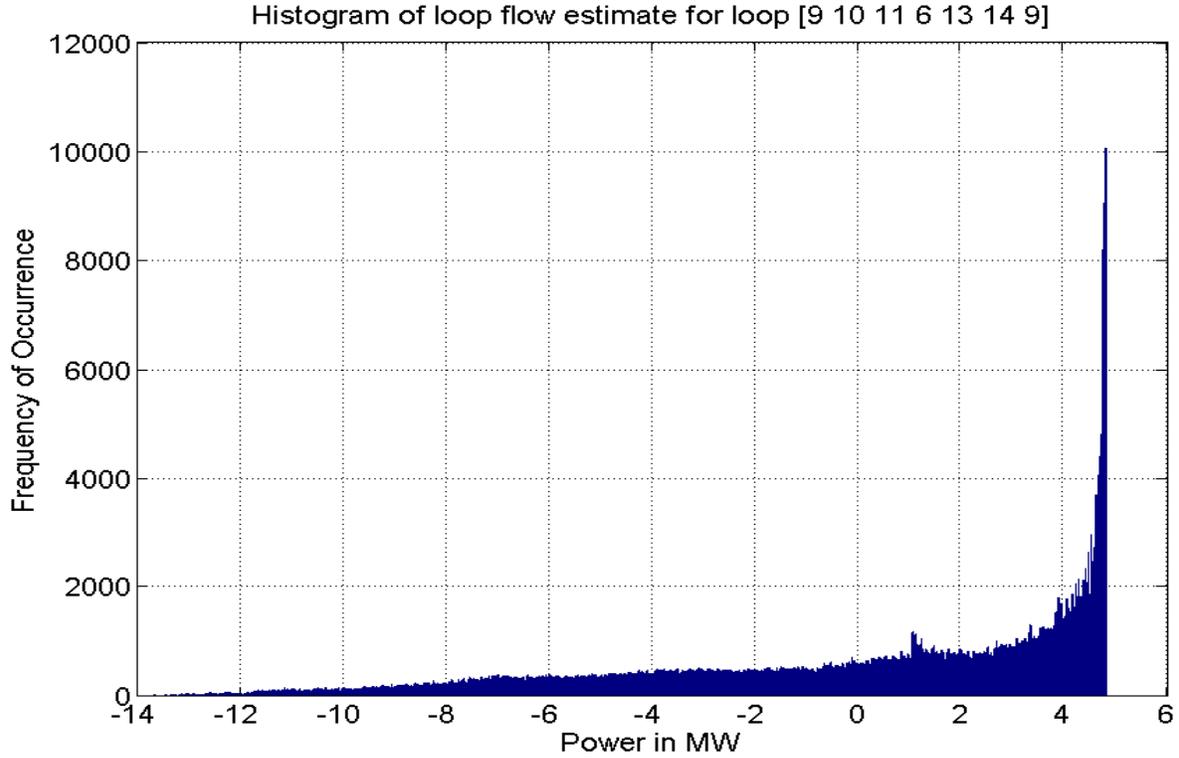
Figure 24 Histogram of estimated loop flow for the *Loop 8* using OLS.

Table 11: Correlation between USFs and Wind Farm Penetrations at Different Buses

Estimated loops flow ($\widehat{x_{ols}}$)	Wind farm output at bus 5	Wind farm output 1 at bus 6	Wind farm 2 output at bus 6	Wind farm output at bus 7	Variance of loop flows
Loop 1	0.92	0.79	0.68	0.59	101.55
Loop 2	0.8989	0.80	0.69	0.63	31.10
Loop 3	-0.936	-0.74	-0.64	-0.67	27.90
Loop 4	0.90	0.77	0.67	0.67	33.09
Loop 5	-0.71	-0.88	-0.74	-0.55	19.85
Loop 6	-0.96	-0.65	-0.57	-0.71	25.73
Loop 7	-0.90	-0.73	-0.66	-0.26	8.23
Loop 8	-0.88	-0.78	-0.68	-0.67	19.86

As seen from Table 11, the output of wind farm at bus 5 has a strong positive correlation

with estimated *Loop 1* whereas a strong negative correlation with the estimated *Loop 3*.

Variability associated with estimated loop flows has to be investigated using different statistical

tools. The value of the correlation index is a function of both the degree of connectivity of the bus and the market expected flows on lines connected to that bus. The largest annual variance is observed for the estimated loop flow associated with *Loop 1* due to the large variance of wind farm output at bus 5. Additionally, the installed capacity of this wind farm is significantly larger than the others, thus imparting comparatively greater variability to the estimated loop flow. Similar inferences can be drawn for the rest of the estimated loop flows and wind farm outputs. The location of the wind farms is crucial in this study such that the inferences drawn from the estimated values will change significantly, if connected to different buses. This concludes the demonstration of the application of GIS based synthesis of system matrix to accommodate USFs.

3.7   Concluding remarks

An application of a GIS technique in estimating minor loops in an electric network is explored. Simple linear estimator is used to estimate the minor loop flows using the system information and measurements of unscheduled flows. Line layouts within loops are synthesized using the GIS coordinates of the buses to automate the formation of the system (incidence) matrix. USFs on transmission lines (branches) are accommodated using the estimates of the loop flows, which in this case are estimated using ordinary least squares. The system matrix for the IEEE 14-bus test system was synthesized to be used to estimate loop flows in an annual variable generation scenario. Loop flow estimates show positive and negative correlations with wind farm output depending upon the remoteness of loops from the wind farm bus, but no fixed pattern was discernible.

CHAPTER 4

SOLVING THE LINEAR ESTIMATOR

Chapter 4 explains the structural deficiency of the incidence matrix and solutions in the linear estimator. The incidence matrix synthesized for any test network using the automated techniques discussed in the chapters 2 and 3 may have numerical ill-conditioning. The common inference in such incidence matrices with numerical ill-conditioning is the multicollinearity of regressors. OLS technique as applied in the chapter 3 to estimate loop flows does not provide the best possible estimates in such cases. RR uses a bias parameter to minimize the average squared error, instead of the ordinary least squares method, when numerically ill-conditioned system matrices are present. RR has been proposed to counter multicollinearity; however application to bulk interconnections requires visual decision making. Sections 4.2 and 4.3 demonstrate an analytic technique of choosing a bias for RR. No human intervention of gauging ridge traces are needed in the proposed approach and hence can be readily applied to a large size of regressors (i.e., loop flows) to be estimated. An application of regression model building and robust regression to accommodate unscheduled flows (USFs) is presented as an alternative to analytic RR, for numerically ill-conditioned incidence matrices in sections 4.4 and 4.5. Constraints for model reduction are developed on the basis of the statistical significance of regressors and power system operations. Robust regression (M-estimators) is applied to solve the linear estimator and counter the impacts of potential outliers, if any. Applications of analytic RR and building the regression model are demonstrated on the IEEE-14 bus test system. This chapter has been directly taken from the references [56] and [63]. Reference [63] is under review at the time of writing this dissertation.

4.1   Introduction

Unscheduled flows (USFs) are the deviation of transmission line flows from expected levels due to inconsistency between physical laws of electricity and electricity trades [29]. A linear estimator to model USFs using a mathematical artifact termed 'minor loop flows' shown in equation (2) has been proposed. The minor loop flow vector, $\hat{x}_{ols}$, is estimated using the left pseudoinverse of incidence matrix shown in equation (3). The set of loops selected for the synthesis of incidence matrix in (2) accounts for all the transmission lines in the network that experience USFs. Loop selections can be performed by two techniques: a) visual synthesis, and b) algorithmic synthesis, suitable for small networks with precise visuals, and practical bulk interconnections respectively [35]. Visual synthesis derives loops by human decision-making using available one-line diagrams. Algorithmic synthesis relies on mathematical techniques like graph theory and does not require visuals. The philosophy of synthesis of the incidence matrix is based strictly on accommodating the USFs. Hence, this may lead to an overrepresented set of equations that implies the regressors in (2) may not be perfectly orthogonal.

Least squares minimization yields the best solution only when the incidence matrix is numerically well conditioned. In most cases, the incidence matrix is ill-conditioned since it is formed heuristically without any eigenvalue constraints such as in [29]. The condition number of $H^T H$ is significantly large indicating numerical ill-conditioning of the incidence matrix. Large condition number and other related statistical measures indicate near linear dependence of the regressors [29]. The estimates of loop flows with multicollinear incidence matrix have large variances and thus are numerically unstable. In such scenarios, the unstable estimates provided by equation (3) are less reliable. Also, if the estimates are intended for use in a compensatory pricing mechanism as in [29], then robust methods that yield estimates of high confidence, such

as the RR, are essential. The OLS solution shown in (3) is applicable when the regressors are at least near orthogonal. A solution to overcome multicollinearity of incidence matrix and obtain better quality estimates than OLS, by using analytic RR is proposed in sections 4.2 and 4.3 [56].

The minor loop flow estimation framework given in [29] is for deterministic generation and load values. Future electricity grids are expected to have variable generation such as wind farms connected at the transmission level. The stochasticity in wind is expected to alter the USFs in the system. Specifically, wind farms will impart variability and uncertainty to USFs. A demonstration of the estimation of minor loop flows used to accommodate USFs in high wind energy penetration scenarios is presented here. Analytical RR is chosen as the robust estimation technique as it provides the value of bias directly, regardless of the size and characteristics of the network, and by analytical decision-making. Some metrics for quantifying the improvement of the minor loop flow estimates by RR over ordinary least squares (OLS) methods are given.

An alternative technique based on stepwise model building is proposed to counter multicollinearity of the incidence matrix. Accurate estimates of loop flows are highly recommended as they may be used in allocating USFs responsibility to individual players, and hence assist in market accommodation.

4.2   Ridge regression

To overcome the effects of multicollinearity, RR introduces a biasing factor k in equation (3), which can be chosen either graphically or by analytical selection. In the former, ridge traces of the regressors, $\hat{x}_{rr}$, are plotted against k, which varies from 0 to a reasonable upper limit. Ridge traces provide a graphical insight into the stability of estimates for different k values. A suitable k value is chosen for a stable set of regressors based on inspection. Such a procedure,

subject to human decision-making, may yield varied and suboptimal k values; and, for large number of regressors, visual selection becomes cumbersome. Analytical selection is a more effective method as it provides a suitable value of bias by accounting for the OLS estimates. Point estimate of regressor variances, σ_{per}^2, for the biased estimator is smaller as compared to that of the unbiased estimator with multicollinear incidence matrix [60]. Reduction in σ_{per}^2 of regressors indicates stability in regression. For RR, σ_{per}^2 is the sum of variances of parameters in $\hat{x}_{rr}$ and the square of bias, hence larger value of k leads to smaller σ_{per} and larger bias [60]. Equations (4)-(6) depict the ridge estimates, the bias value, and σ_{per}^2 of regressors for the proposed estimator respectively, [60], [76],

$$\hat{x}_{rr} = (H^T H + kI)^{-1} H^T z \tag{4}$$

$$k = \frac{p\sigma_{ols}^2}{\hat{x}_{ols}{}^T \hat{x}_{ols}} \tag{5}$$

$$\sigma_{per}^2(\hat{x}_{rr}) = \sigma_{rr}^2 + (bias\ in\ \hat{x}_{rr})^2, \tag{6}$$

where, $\hat{x}_{rr}$ is the ridge regression estimate, p is the total number of loops, σ_{ols}^2 and σ_{rr}^2 are the variance of estimates obtained using OLS and RR, respectively. The biasing factor k increases the spread of the eigenvalues of incidence matrix and hence improves the quality of estimates [76].

4.3   A case study

The IEEE 14 bus test system, given in [51] is modified by adding four wind farms with regulated voltages (one each at buses 5 and 7, and two at bus 6). This modified system is used to simulate the probabilistic load flow based on the Monte Carlo (MC) technique. Eight loops in the

test network, shown in Table 12, are chosen visually for accommodating the USFs [35]. Wind farm outputs are the power injections by the respective farms into the grid measured at the point of common coupling. The objective of this simulation is to observe the distribution of USF due to the uncorrelated wind farm and load variations with respect to a single base case. The load values are randomly chosen from standard distributions obtained directly from an AC probabilistic load flow analysis [51]. The schedules (in MW) corresponding to the buses for the base case in the format {bus no., gen, load} are $\{1, 186.9_{cg}, 0_l\}$; $\{2, 40.04_{cg}, 21.74_l\}$; $\{3, 0, 94.2_l\}$; $\{4, 0, 47.8_l\}$; $\{5, 20.73_{wg}, 7.6_l\}$; $\{6, 18.03_{wg}, 11.2_l\}$; $\{7, 4.48_{wg}, 0_l\}$; $\{8, 0, 0_l\}$; $\{9, 0, 29.5_l\}$; $\{10, 0, 9_l\}$; $\{11, 0, 3.5_l\}$; $\{12, 0, 6.1_l\}$; $\{13, 0, 13.5_l\}$; and $\{14, 0, 14.9_l\}$, where the subscripts *cg*, *wg*, and *l* denote conventional generator, wind-powered generator, and load, respectively. The transmission line flows for the base case are obtained from converged power flows using the scheduled values listed. Base case flows are assumed to be the market expectations as they represent the most likely scenario used to calculate *z* in (4).

To account for the statistical variance of the wind farm output and the loads, approximately 500000 iterations are performed using the MC technique to obtain the transmission line flows. The minor loop flows estimated by OLS is used in 5) to obtain the appropriate *k* for the revised estimation provided in (4). For all the eight regressors (estimated minor loop flows) in this case study, reduced variance was observed when using the analytical RR compared to OLS as shown in Table 12. The accuracy of RR in this analysis is indicated by: reduction in average squared error by 2.76%; and reduction in variance of squared error distribution by 54% [76]. The externally studentized residuals of the OLS and RR estimates conform to normal distribution. The maximum values of the adjusted coefficient of multiple determination (adjusted R^2) of the entire dataset (i.e., approximately 500000 iterations) of the

OLS and RR fits are 0.83 and 0.85, respectively. These two observations indicate the accuracy of the approach [60]. The application of the proposed technique to bulk interconnection data will involve the use of large datasets for branch flows and incidence matrices representative of the network size. Also, an algorithmic synthesis of loops becomes imperative, as bulk interconnections are multi-planar with complex connections and are difficult to process visually. The disadvantage of the RR technique is the requirement of the OLS estimates for calculating the biasing parameter k.

Table 12: Comparison of Minor Loop Flow Estimates in the IEEE-14 Bus Test System Using RR and OLS Methods

Loop number	Nodes of the system in the loop	Statistics of estimates by OLS		Statistics of estimates by RR	
		Mean (MW)	Variance (MW^2)	Mean (MW)	Variance (MW^2)
Loop 1	[2 1 5 2]	-0.13	185.2	0.33	79.30
Loop 2	[2 3 4 2]	-0.06	66.97	-0.02	40.04
Loop 3	[6 12 13 6]	0.01	49.80	-0.18	13.24
Loop 4	[4 3 2 5 4]	-0.13	92.75	-0.03	51.61
Loop 5	[4 5 6 11 10 9 4]	0.12	51.99	-0.05	17.88
Loop 6	[4 7 9 14 13 12 6 5 1 2 4]	0.01	40.23	-0.35	21.00
Loop 7	[4 9 7 4]	0.04	17.66	0.09	3.13
Loop 8	[9 10 11 6 13 14 9]	0.13	64.71	0.09	18.83

4.4   Building the regression model

USF accommodation using the linear estimator is a *controlled experiment* as the modeler or the analyst controls the regressor selection. Building the regression model involves four steps – data collection; regressor selection; refinement; and, validation [57]. Data collection is relatively easier as it is procured either from commercially available power flow software or actual measurements of real power flows in the network. Selection of regressors can be governed by their respective statistical significance. However, regressors for loop flow estimation are based on two sets of information — set of loops and layout information of transmission lines. The number of regressors should adequately represent the network to accommodate USFs.

Model reduction (MR) reduces the number of regressors used in synthesizing the incidence matrix for countering multicollinearity. The incidence matrix synthesized with all possible closed loops is the *full* model; whereas the *H* synthesized with any fewer regressors is the *reduced* model. Any model is treated as provisional and is subject to model refinement. MR reduces the regressors subject to power systems constraints and statistical significance. Critical lines that are essential for maintaining network reliability and are historically prone to USF are preserved in any model. The constraints to reduce the number of regressors in the incidence matrix are: (a) no regressors corresponding to a critical line can be removed; and (b) no two regressors that are used in representing a line can be removed. Constraint (a) ensures retention of all information related to critical lines; constraint (b) ensures that no transmission line is completely excluded from the *reduced* model.

Complying with the above-mentioned constraints, the following procedure to obtain the *reduced* incidence matrix is proposed. The rank of the *full* incidence matrix provides the total number of regressors required to sufficiently represent the system. The decision to remove

regressors is based on the t-statistic (t-stat) [58]. A formal hypothesis test is set for individual regressors with the null hypothesis: $\hat{x}_k = 0$, and an alternate hypothesis: $\hat{x}_k \neq 0$, for computing t-stat and its probability (p-value) at a significance level (usually 5%) [58]. The structural stability and multicollinearity of the *reduced* model is assessed using standard diagnostics [96]. However, model validation and assessment are performed for each potential model. Potential outliers in the observations and its influence on regression plane are gauged using Cooks distance [97]. For the *full* and the *reduced* model, robust estimates with different weight functions (w) — Bisquare (*BSQ*), Welsch (*W*), and Fair (*F*) — are obtained to counter the effect of potential outliers. The objective function of robust regression, shown in equation (7), is minimized using all the three weight functions ($w(r)$), where r is the vector of residuals. The tuning constants for the weight functions of *BSQ*, *W*, and *F* are 4.685, 2.985, and 1.4, respectively as shown in Table 13 [59]. The above-mentioned weight functions are chosen on account of the diverse values of the respective tuning constants.

$$w(z - Hx)x = 0 \tag{7}$$

Table 13: BSQ, W, and F Weight Functions [59]

Name	Weight Function	Range
Bi-squared	$(1 - (r/4.685)^2)^2$	$\lvert r \rvert \leq 4.685$
Welsch	$e^{(-(r/2.985)^2)}$	—
Fair	$(1 + \lvert r \rvert/1.4)^{-1}$	—

4.5   A case study

A probabilistic load flow on the basis of the Monte Carlo technique to accommodate the variability of wind is simulated on the IEEE-14 bus test system according to [56]. Transmission lines 1-2 and 1-5 are identified as critical lines, as the slack bus is connected at bus 1. The objective of this simulation study is to obtain the *z* vector in equation (3) for an annual scenario of wind power and known market conditions. The total number of loops in the full model is eight [56]; however, the rank of *H* is seven, indicating a possibility of multicollinearity. Loop 3, on account of its insignificant t-stat (p-value > 0.05) and lack of any critical lines in the path, is excluded to form a reduced model. The multicollinearity index — the base-10 order of magnitude of the condition number of the incidence matrix —decreases from 16.65 for the *full* model to 1.50 for the *reduced* model.

For an OLS-based loop flow estimation using the *full* model, the potential outliers measured by the Cooks distance are observed for lines 2-3, 3-4, and 4-7 [97]. The impact of the outliers on the regression plane depends on its respective location. Table 14 shows the estimates for both the *full* and the *reduced* models by robust regression techniques. Some linear fits exhibit negative coefficients of multiple determination (adjusted R^2), which is frequent in cases of linear estimators without intercepts. Since *minor loop flows* do not exist in the absence of USFs, intercepts do not exist in this experiment. In such cases, the adjusted R^2 has no clear interpretation [98]. The adequacy and validity of the model is checked by standard techniques such as normal probability plots, scatter plots. The regression model forces Loop 3 to be zero on account of its insignificant value of t-stat. In the *reduced* model the regressor corresponding to Loop 3 is deleted from the estimation procedure at the onset. No uniform increase or decrease in the variance of loop flow estimates is observed for any particular weight function. A comparison

of the statistics of the loop flow estimates from the *full* and the *reduced* model indicates that MR leads to decrease in the variances of a few regressors (loops 1 and 2) and a significant increase in the variances in others (loops 5 through 8). The order of the largest variance inflation factor, $max(\log_{10}(VIF))$, decreases from 16.54 for the *full* model to 0.68 for the *reduced* model, thus indicating a trend toward greater confidence in the estimates obtained using the *reduced* model [55]. Note that lower values of the tuning constant provide better estimates of loop flows with lower variances. The normal probability plots of the residuals from *reduced* models fits reveal no potential outliers.

4.6 Comparison of loop flow estimates for the IEEE-14 bus test system

In this chapter, loop flow estimates for the IEEE-14 bus test system are obtained using three techniques namely, OLS, analytic RR, and robust regression. Robust regression is applied to the full and reduced model to obtain the loop flow estimates of eight and seven loops respectively. The quality of estimates can be interpreted by the values of mean and variances of the respective distributions obtained from the various techniques. The intent of any generation scheduling technique may strive to supply loads at an optimum cost as well as keep USF at manageable levels. However, there is no formal proof to support USF minimization as a scheduling objective. Unavailability of measurement of loop flows (considering that loop flows are mathematical artifacts) renders it difficult to perform accuracy checks on loop flow estimates. The estimation of loop flows is statistically ensured to be within reasonable accuracy and hence can be compared. The selection of the estimation technique is based on the objectives of loop flow estimates and accommodation technique. A simplified decision tree on the method of estimation to be adopted is shown in Figure 25:
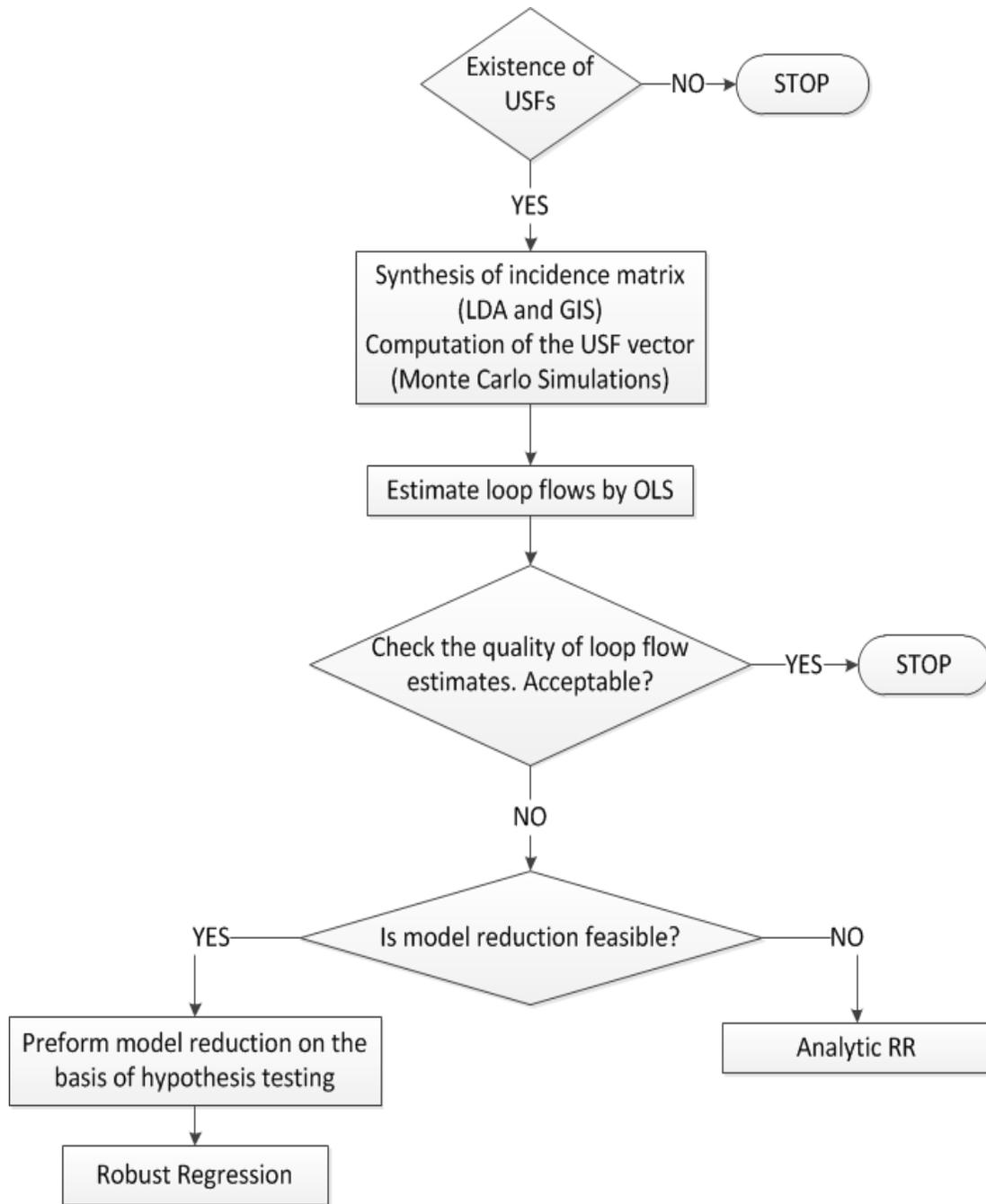
Figure 25: Decision tree for adopting suitable estimation technique for modeling USF by using loop flows

The loop flow estimates obtained by both OLS and analytic RR have mean values close to zero; however they have dissimilar variances. The values of variances of all eight loop flows in the IEEE-14 bus test system, as mentioned earlier, demonstrate significant reduction when

analytic RR is used instead of OLS. Hence, the loop flow estimates from OLS and analytic RR are solely compared on the basis of variance. Reduction in variance implies an improvement in estimation as loop flow estimates are more certain. The iteratively reweighted least squares (robust regression) with different weighting functions (Fair, Welsch, and Bisquare) are used to counter the impact of potential outliers as mentioned in section 4.5. This technique estimates the loop flow 3 as zero for all the iterations of the MC simulation with the full model. Thus, the variance of loop flow 3 is zero as well. In practice, the estimate of any loop flow cannot be zero due to the interconnected nature of networks and physical laws of electricity that govern power flow. Mean values of distributions of loop flows (2, 4, and 7) are estimated to be significantly greater than zero as compared to OLS and analytic RR statistics. This is a noteworthy outcome arising from the fact that loop flow 3 has been estimated as zero leading to an increased mean in other regressors. Variance of loop flows 1 and 2 demonstrate a reduction from full to reduced model, whereas loop flows 5 through 8 indicate an increase in variance. The variance of loop flow 4 shows no change in variance. Similarly, trend of changes in the mean and variance of distributions of estimated loop flows is observed for the reduced model as well. This increase in the mean and variance may be attributed to the fact that all eight regressors are collectively important towards estimation. In other words, loop flow 3 may not be significant individually but provides information for estimation collectively. Further investigation is required in the form of statistical hypothesis testing to confirm the importance of individual regressors in a collective sense.

4.7   Conclusion

The analytical RR is shown as a robust alternative to the ordinary least squares based estimation of minor loop flows. Significant improvement in estimates obtained from multicollinear system matrices, an analytical procedure, and ease of implementation may render this technique useful in practical problems such as the accommodation of USF in bulk interconnections. Model reduction — a type of regression modeling technique — based on statistical significance is demonstrated as a solution to combat the effects of multicollinearity of incidence matrices in the accommodation of USF. An application of robust regression to counter the impact of potential outliers is also presented.

Table 14: Comparison of Minor Loop Flow Estimates in the IEEE-14 Bus Test System Using Robust Regression

Loop number	Nodes of the system in the loop	Statistics of estimates by F				Statistics of estimates by W				Statistics of estimates by BSQ			
		Full ($p=8$)		Reduced ($p=7$)		Full ($p=8$)		Reduced ($p=7$)		Full ($p=8$)		Reduced ($p=7$)	
		μ (MW)	σ^2 (MW^2)	μ (MW)	σ^2 (MW^2)	μ (MW)	σ^2 (MW^2)	μ (MW)	σ^2 (MW^2)	μ (MW)	σ^2 (MW^2)	μ (MW)	σ^2 (MW^2)
1	[2 1 5 2]	2.27	285.85	19.20	18.40	0.95	306.82	18.20	19.44	0.84	309.50	18.11	20.05
2	[2 3 4 2]	-18.07	106.88	-1.14	37.63	-18.36	114.59	-1.12	52.49	-18.42	116.54	-1.15	58.06
3	[6 12 13 6]	0	0	–	–	0	0	–	–	0	0	–	–
4	[4 3 2 5 4]	-16.92	223.03	16.92	223.03	-17.24	237.88	17.24	237.88	-17.27	240.39	17.27	240.39
5	[4 5 6 11 10 9 4]	-4.86	40.80	12.06	150.25	-5.23	41.33	12.01	158.75	-5.27	42.10	12.00	159.83
6	[4 7 9 14 13 12 6 5 1 2 4]	3.01	11.03	19.93	193.16	2.87	10.44	20.11	212.99	2.86	10.43	20.14	216.36
7	[4 9 7 4]	18.51	20.63	35.44	165.47	17.00	38.27	34.25	156.56	16.77	45.98	34.05	158.46
8	[9 10 11 6 13 14 9]	-2.51	52.87	14.40	122.11	-2.74	55.02	14.50	129.06	-2.76	55.53	14.51	130.41

CHAPTER 5

FORECASTING ERRORS IN PREDICTED WIND POWER AND USF

This chapter documents the analysis of impacts of forecasting errors associated with the power output of wind power plants for an annual systems-planning case. The prediction intervals instead of accurate point forecast values of power output of wind plants have been computed from normal distribution. The normal distribution has been found in past literature to be suitable for modeling forecasting errors in predicting output of wind plants for an annual horizon of forecast. The structure of analysis will remain exactly same if a USF scenario for a contract period is to be simulated and investigated. The choice of forecasting error distribution of wind power may depend on the time frame of analysis. Model adequacy and statistical inferences of the loop flow estimates is discussed. Impact of forecasting error on distributions of estimated loop flow is explored on the basis of Kolmogorov-Smirnov (KS) and chi-square goodness-of-fit tests. This chapter has been directly created from the authors publication referenced - [44].

5.1 Introduction

USFs exist in power systems at transmission level on account of the interconnections and deregulation of network. USFs are defined as the difference between market-expected line flows and actual line flows [29]. Interconnections provide multiple paths for power flows in a network. In a deregulated market, financial transmission rights (FTRs) are allocated to electricity trades based on equitable distribution of transmission infrastructure as per the FERC directives [82]. Market mechanism of assuming flows on transmission lines using multiple electronic tags is based on the superposition theorem. This leads to an inconsistency between actual and expected flows [29]. USFs lead to: a) forced participation of transmission infrastructure; deviation from the already cleared market prices; c) congestion in lines leading to near steady state operation; and, reduction in available transfer capacity (ATC) [29]. In this regard, it is required to model and accommodate USFs.

USFs may be accommodated technically, [29], or in the market [5], [14], [15], [17], [20], [71], [73]. Technical accommodation methods involve managing USFs at the operational level such as rerouting power flow using flow control devices and curtailment of schedules at certain times. This technique is effective when the security constraints of the network are not violated maintaining network reliability. Power and energy supply indices enforced by regional and local entities have to be satisfied. Currently, the Eastern and the Western Interconnections in the US practice distinct methods of USF accommodation [20].

The market accommodation method is a post gate closure method that requires a systematic analysis of the inconsistency between the expected and actual power flows on the branches. Several studies to develop improved economic models to charge or compensate utilities for USFs based on power transfer distribution factor (PTDF) and its modifications have

99

been developed [5], [15], [73]. Additionally, a contribution factor based on the estimation of a mathematical artifact *called minor loop flow*, has been developed for use in a pay/compensate mechanism for GENCOs [29]. In this method, the individual impact of GENCOs on USFs is calculated using the estimated *minor loop flows* from the respective e-tags (schedules) and the *aggregate schedule*.

Worldwide, numerous studies on the impact of USFs specific to operating grids such as New Zealand, Europe, North America, and south China have been undertaken [13], [14], [17], [20], [71], [99]. Significant wind power penetration at the transmission network is a growing trend in the U.S. and is expected to continue. The U.S. Department of Energy has a target of supplying 20% electricity demand by wind generated power [1]. Wind farm outputs are subject to variability and uncertainty. Hence, power output scheduling is influenced by forecasting models and error statistics. Forecasting error distribution is governed by the timescale of the forecast and the type of forecasting model itself [41]. Long-term forecasts are usually more prone to error as compared to short-term forecasts [100]. Hence, future trade commitments by the wind power generation companies (WGENCO) might be subjected to greater uncertainty. WGENCOs have been found to operate with greater economic efficiency in markets with shorter time resolution of closure [101]. Another factor contributing to the accuracy of forecast models is the geographical span of regimes. Individual wind farm output forecasts are found to be less reliable than collective wind farm output forecasts since smoothing of errors occurs [39]. The economic impact of forecast error costs may be as high as ten percent of the total wind producer income [101].

In this paper, an analysis of the impact of forecast error distribution of WGENCO output on the accommodation of USFs, and related statistical inferences are presented. Sections 5.2 and

5.3 discuss the linear estimation framework and the statistical processing of wind farm output and errors respectively. Section 5.4 describes the IEEE 14 bus test system with wind farms and the loop flow estimation results. Section 5.5 examines the quality of loop flow estimates obtained and section 5.6 concludes.

5.2   Estimation of loop flows

USFs are accommodated using a mathematical artifact termed as *minor loop flows* using ordinary least squares (OLS) linear estimator, shown in (2), and (3), respectively [29]. However, the system matrix may exhibit multicollinearity that can make the OLS estimate unreliable [57]. A robust estimation technique such as the analytic ridge regression (RR) may be adopted to counter the ill conditioning of the system matrix. RR involves biasing the linear estimator with a suitable non-negative value to improve the quality of estimates. (4) provides the solution by ridge regression and 5) is the analytical formula for calculating the bias value, k [56], [57].

5.3   Wind power forecasting errors

WGENCO commitments for the day-ahead market are influenced by short-term forecasting errors [101]. For the point estimates of the wind farm outputs the interval estimates are obtained using the normal distribution modeling the forecast error. The forecast error distribution is modeled as $N(0, var(P_t))$, where $var(P_t)$ is the variance of WGENCO output data with time step t [40]. The time step is dependent upon the required forecast horizon and desired accuracy of forecasting. Typical mean absolute errors (MAE) for system-wide wind power forecast models for day-ahead and hour-ahead cases are in the range of $15 - 18\%$ and $6 - 11\%$, respectively [100]. The typical MAE is mapped onto the error distribution to determine the region of the most expected errors as given in equation (8):

$$Forecast\ mean\ absolute\ error\ =\ F_{\in 1}\ <\ \in\ \leq F_{\in 2} \tag{8}$$

where $F_\in$ is the cumulative distribution function (CDF) of the forecast error distribution with an appropriate time step and $\in$ is the error variable. The value of expected error will be from either of the two aforementioned typical MAE intervals based on market resolution and prediction period. The accurate point estimate corresponds to no error and hence the typical error area intervals lying around it. The two calculated error values, $\in_1$ and $\in_2$, are used to obtain the interval forecasts. Given the nature of the error distribution, the values of $\in_1$ and $\in_2$ are approximately equal in magnitude and opposite in sign. Interval forecast error centered on zero is chosen for this analysis to obtain the interval forecast. Similar mapping can be done elsewhere in the error distribution to obtain error intervals, however they are less likely to occur than the one chosen. Each value of accurate point estimate of x may be forecasted as any value in the interval $[(x + \in_1), (x + \in_2)]$. Smaller forecast horizon values imply smaller magnitudes of $\in_1$ and $\in_2$ and vice versa.

5.4   Test case

Wind farms in the test system

The IEEE 14 bus test network is used to investigate the impacts of annual wind power forecasting errors on USFs. Details of the test system are available in [51]. Four wind farms are connected, one each to buses 5 and 7, and two to bus 6, of the test network. The distances between wind farms are approximately equal to the line lengths. Wind farm power measurements at the point of interconnection are assumed to be the accurate point estimates[†]. Forecasted interval estimate for the wind farm power output for one year is emulated using the accurate point forecast and error interval selection as discussed in section III. Annual accurate point estimate with time resolution of 1 minute is used in equation ((8) to model the forecast error.

102

This error model represents the largest forecast horizon and hence provides the most conservative interval forecast.

Using the annual standard deviation of the accurate point estimates of the individual wind farms, four error distributions are calculated as described in section III. A typical day-ahead MAE of 17% is chosen to further obtain the error interval from the error distributions. Upper and lower forecast bounds of accurate point estimate for all WGENCO outputs are obtained using a positive error correlation between the distributions. The installed capacity, location, the forecasting error distribution, and the error interval of the four wind farms are shown in Table 15. As an example, for a single day (January 1st, 2004), the accurate point estimate along with the lower and upper bounds for wind farm output at bus 5, calculated using the error distribution from Table 15 are shown in Figure 26. The negative values observed in the lower bounds shown in Figure 26 are obtained analytically. For practical purposes, the negative values are accounted as zero power generation by WGENCOs.

Table 15: Location, Installed Capacity (MW), Annual Forecast Error Distribution, and Error Intervals of Wind Farms in Test System

| WGENCO Location	Capacity	Forecast Error Distribution	Error intervals
Bus 5	122.8	N(0, 27.33)	[-6.01, 6.01]
Bus 6	95.9	N(0, 12.62)	[-2.78, 2.78]
Bus 6	99.0	N(0, 8.53)	[-1.88, 1.88]
Bus 7	49.8	N(0, 5.62)	[-1.24, 1.24]
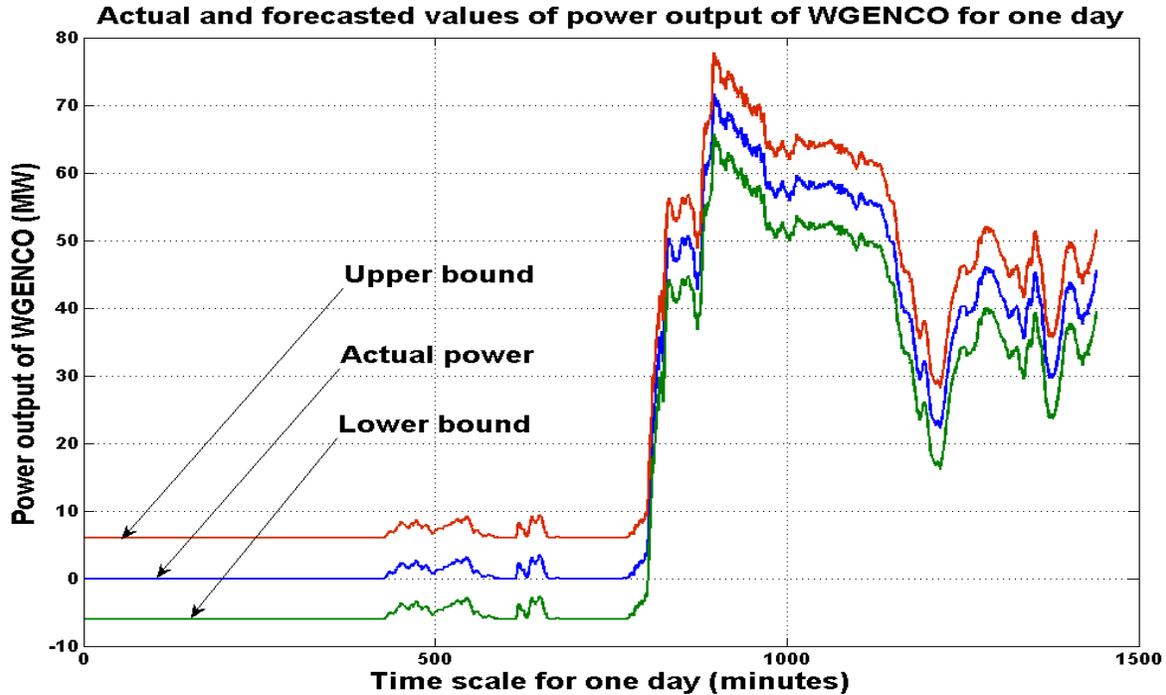
Figure 26 Accurate point estimate and bounds on the power output of the WGENCO at bus 5 for one day.

Loop flow estimation with interval forecasts

Individual schedules for the participating GENCOs are shown in Table 16. The commitment of the WGENCOs is calculated from the average of annual accurate point estimate. Expected flow for the individual schedules is obtained by using the e-tags of the respective GENCOs from Table 16. Superposition of the individual schedules provides the aggregate schedule for the test system [29]. Hence, the expected flows for the aggregate schedule are computed using expected flows of all the individual schedules, shown in Table 17.

Running the power flow algorithm on the individual and aggregate schedules yields the actual line flows. Monte Carlo (MC) simulations with wind farm output as a variable are executed for an annual scenario according to [35]. Accurate point estimate, upper bound and lower bound estimates are the three different wind farm output cases executed independently in

104

the MC simulation. USF vector z is calculated as the difference between the actual and expected line flows. The system matrix H in (1) is synthesized based on a heuristic technique used in [35]. Figure 27 illustrates the connectivity and minor loops in the test network and Figure 28 shows the system matrix mapping the relationship between the minor loops and the branches [35]. Branch 7-8 in Figure 27 has an orphaned node 8, i.e., without a load or source connected to it. Hence, this branch does not play any significant role in the analysis. For estimation purposes, the column in H^T corresponding to the branch 7-8 is deleted since it has all zeroes.

Table 16: Individual Schedules for GENCOs in the IEEE 14 Bus Test System

GENCO location	E-tag index	Commitment (MW)	Market path	Sink
Bus 1	1:1	84 MW	1-2-3	Bus 3
	1:2	23 MW	1-2-4	Bus 4
	1:3	29 MW	1-5-4-9	Bus 9
	1:4	15 MW	1-5-4-9-14	Bus 14
	1:5	9 MW	1-5-4 -9-10	Bus 10
	1:6	6 MW	1-5-6-12	Bus 12
	1:7	7 MW	1-5-6-13	Bus 13
	1:8	4 MW	1-5-6-11	Bus 11
Bus 2	2:1	22 MW	NA	Bus 2
	2:2	8 MW	2-5	Bus 5
	2:3	10 MW	2-3	Bus 3
Bus 5	5:1	21 MW	5-4	Bus 4
Bus 6	6:1	11 MW	NA	Bus 6
	6:2	7 MW	6-13	Bus 13
Bus 7	7:1	4 MW	7-4	Bus 4

Table 17: Expected Line Flows in the IEEE 14 Bus Test System for the Aggregate Schedule

| Line	Branch	Expected flow (MW)	Line	Branch	Expected flow (MW)
1	1-2	107	11	6-11	4
2	1-5	70	12	6-12	6
3	2-3	94	13	6-13	14
4	2-4	23	14	7-8	0
5	2-5	8	15	7-9	0
6	3-4	0	16	9-10	9
7	4-5	-32	17	9-14	15
8	4-7	-4	18	10-11	0
9	4-6	53	19	12-13	0
10	5-9	17	20	14-13	0
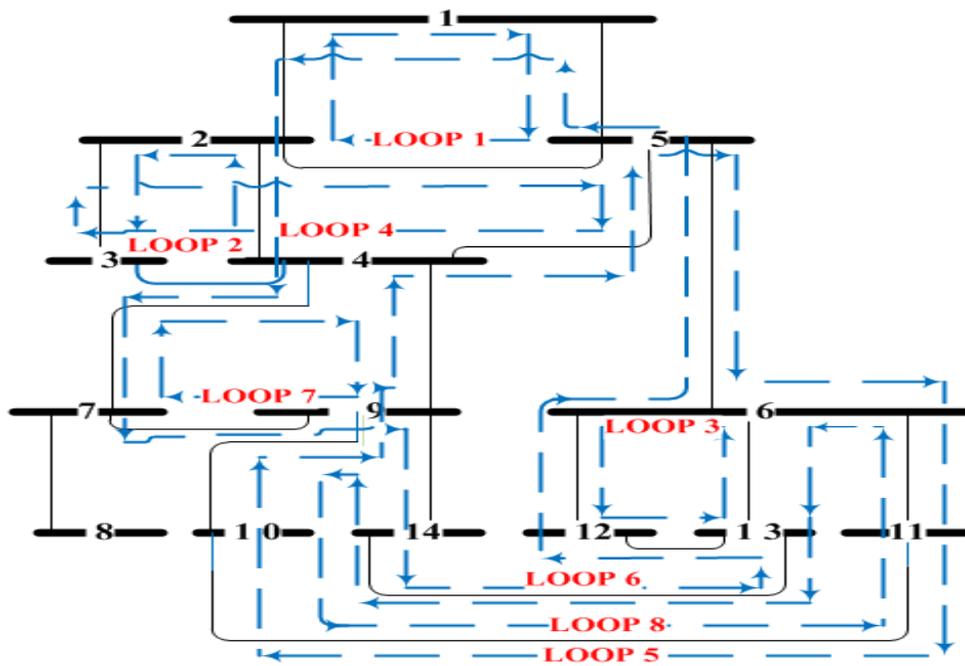
Figure 27 IEEE 14 bus test system showing the 8 chosen loops to synthesize the H matrix [35].

Equations (3), (4), and 5), are used to obtain the RR estimates of loop flows for each of the three wind farm output cases. As an example, Figure 29 shows the histogram plot of the estimated value of the minor loop flow no. 1 - which is bounded by the branches between the following nodes: 2-1-5-2 - with the accurate point estimates, upper and lower bound estimates of WGENCO outputs. Accurate point estimates and both bounds of forecasted power output of all wind farms on a 1 minute time resolution are used to estimate the loop flows and construct Figure 29. Figure 30 shows the plot of a minute-by-minute variation of estimated loop flow no. 1 for 00:00 – 01:00 hour on January 1st 2004 for the aggregate schedule corresponding to accurate point estimates of WGENCO outputs. Both Figure 29 and Figure 30 represent the estimated loop flows obtained from the aggregate schedule. USFs in networks with heavy wind energy penetrations are expected to vary on a minute-by-minute basis. An operational characteristic of the network under such variability should be explored at the lowest possible time resolution.

For a set of power injection from WGENCOs at buses 5, 6 and 7 of 0 MW, 0 MW, and 0.1 MW respectively, the estimated loop flow no. 1 is 0.92 MW. For the forecasted lower and upper bounds on the same set of WGENCO power outputs, the values of estimated loop flow for loop no. 1 are -11.65 MW and -1.26 MW respectively. This is indicative of the uncertainty introduced by the conservative error distributions on the loop flows in the network. Rest of the 7 estimated loop flows exhibit a non-zero value for this operating point of WGENCO outputs and constant load. Analogous mapping of estimated loop flows and forecasted bounds on WGENCO output may provide loop flow trends. Thus, the three histograms in Figure 29 provide the bounds on the maximum variation on estimated loop flows due to forecasting error of WGENCO output.

For individual schedules, the flows obtained from converged power flow can be used in the linear estimation model to analyze the individual WGENCO contributions to the USF

scenario. Similar scenarios of estimated loop flows exist for individual schedules of WGENCOs as well. Figure 31 displays the estimated loop flow no. 1 for the individual schedule with accurate point estimate, lower and upper bound estimates of WGENCO at bus 5. Figure 32 shows the minute-by-minute variation of estimated loop flow no. 1 for 00:00 – 01:00 hour on January 1st 2004 for the individual schedule of WGENCO at bus 5. Highly variable estimated loop flows are observed in minute-by-minute variation plots as shown in both Figure 30 and Figure 32. Comparable patterns of estimated loop flow variations are observed for the remaining individual schedules of WGENCOs in the network.

$$
H^{T} = \begin{bmatrix}
1 & 1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \\
-1 & -1 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & -1 & -1 & 1 & 0 & -1
\end{bmatrix}
$$

Figure 28 The system matrix for the IEEE 14 bus test system. Columns indicate line nos. (1-20) and rows indicate loop flows (1-8).
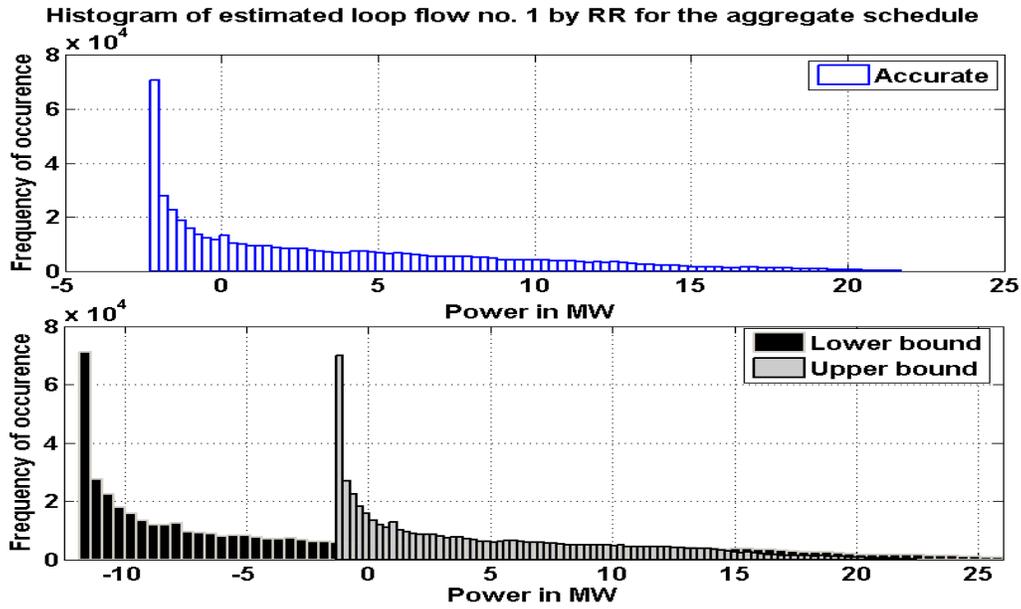
Figure 29 Histogram of the RR estimates of loop flow no. 1 with WGENCO outputs corresponding to accurate point values and forecasted bounds for the aggregate schedule
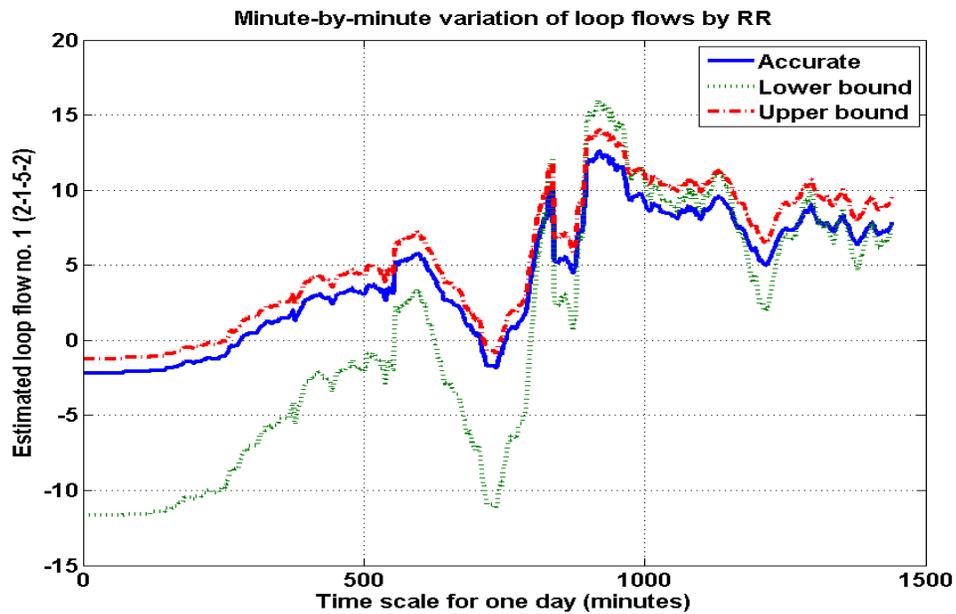
Figure 30 Minute-by-minute variation of the RR estimates of loop flow no. 1 with WGENCO outputs corresponding to accurate point values and forecasted bounds for the aggregate schedule.
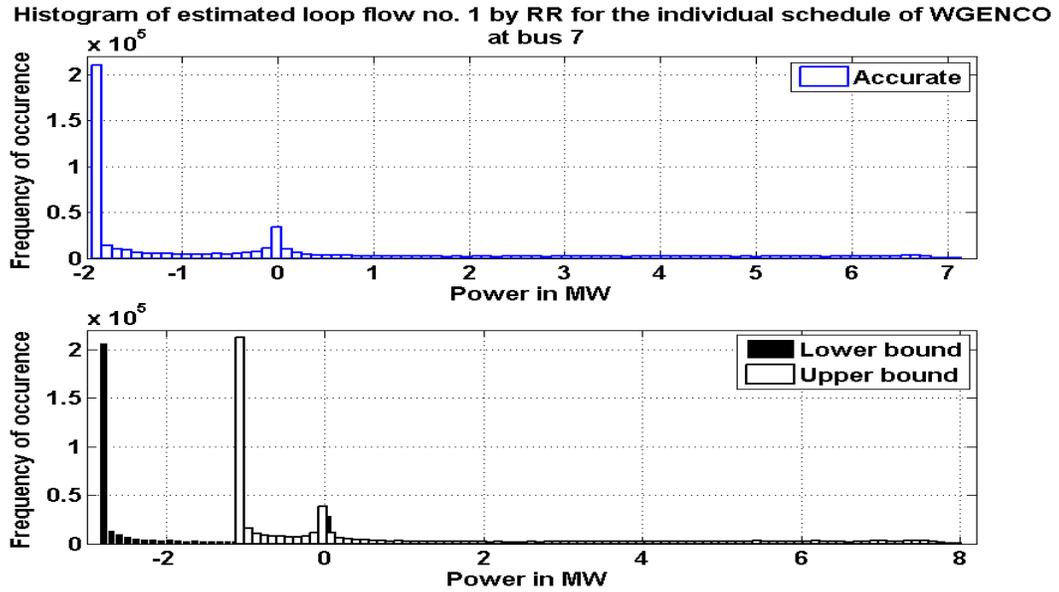
Figure 31 Histogram of the RR estimates of loop flow no. 1 with WGENCO outputs corresponding to accurate point values and forecasted bounds for the individual schedule of WGENCO at bus 5.
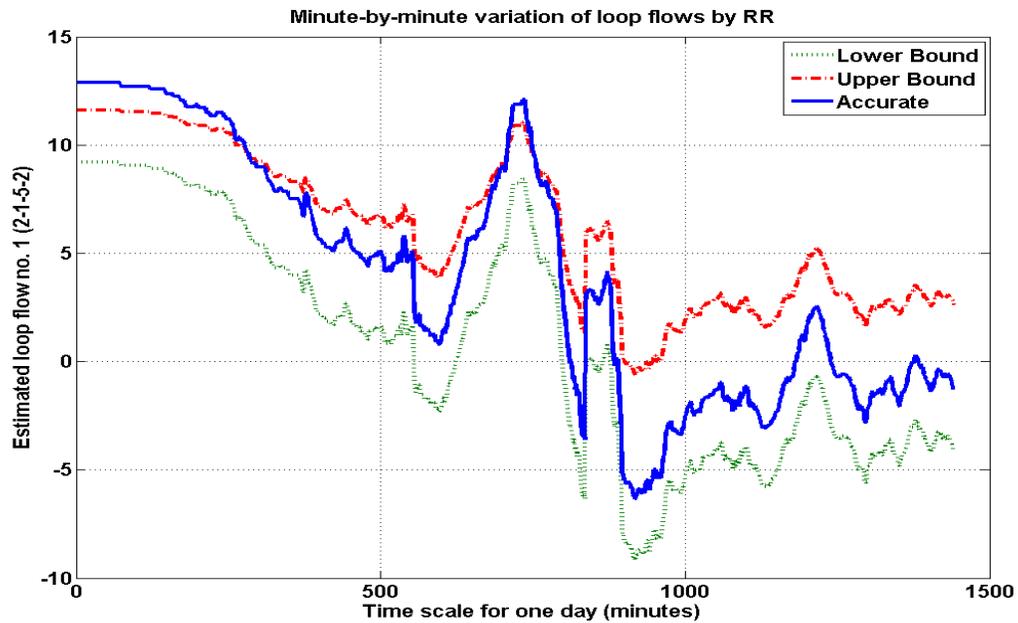
Figure 32: Minute-by-minute variation of the RR estimates of loop flow no. 1 with WGENCO outputs corresponding to accurate point values and forecasted bounds for the individual schedule of WGENCO at bus 5.

110

5.5   Statistical inferences of variable estimated loop flows

By considering the statistical variability associated with the power injections of wind farms in the network, concomitant variability in estimated loop flows in observed as mentioned in the previous section. Statistical inferences from the loop flow estimates obtained are presented in this section.

Normal probability plots of residuals

Residuals of a linear model represent the error between the fitted and actual responses. Normal probability plots (norm plots) provide the information regarding the residuals adherence to a normal distribution. These inferences can be drawn depending upon the location of the data points with respect to the straight line representing a cumulative normal distribution [57]. In an ideal situation all the data points in the norm plots would lie on this straight line. Residuals used for creating norm plots can be calculated in multiple ways. Perhaps the simplest way to calculate residuals for the linear fit is $|H\hat{x} - z|$, for this analysis. Scaling of ordinary residuals assists in isolating the impacts of potential outliers. To study the potential outliers studentized residuals are chosen in this analysis [57]. Figure 33 and Figure 34 show the ordinary and studentized residual norm plots of aggregate schedule and individual schedule of WGENCO at bus 5 as a representative case respectively. Note that there are 19 each of the ordinary and studentized residuals representing the fit of the estimated loop flows in describing the USF on the 19 branches used in the estimation. The reader is reminded of the elimination of branch 7-8 from the analysis.
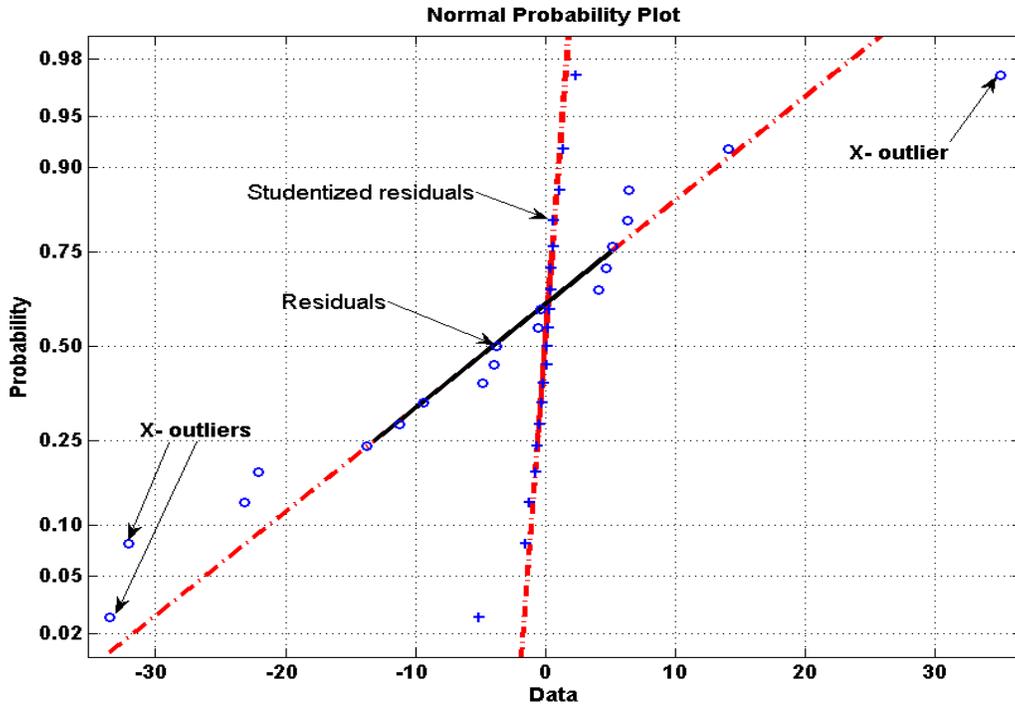
Figure 33 Norm plot using ordinary and studentized residuals for the aggregate schedule indicating adherence to normality of residuals.
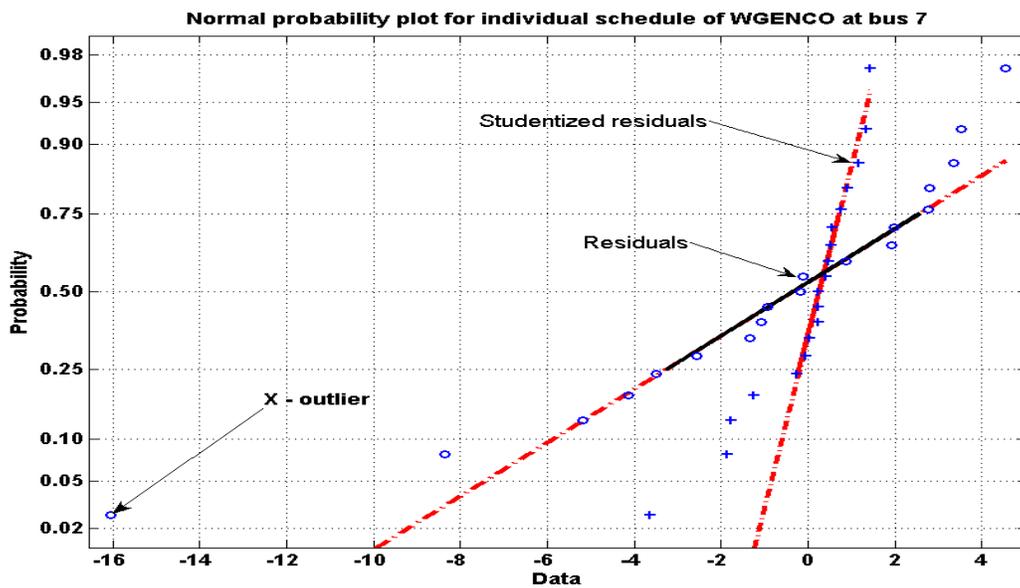

Figure 34 Norm plot using ordinary and externally studentized residuals for individual schedule of WGENCO at bus 5 indicating adherence to normality of residuals.

Both the norm plots for the ordinary residuals exhibit the presence of potential X-outliers. However, the studentized residuals, which are scaled by the respective mean squared errors, reveal that these points lie within an acceptable deviation range from the ideal straight line. To infer, both the studentized residual norm plots can be characterized as light-tailed distributions. Similar inferences may be drawn for other iterations within the MC simulation. These inferences are specific to the representative test system, power injection data, and the set of estimated loop flows shown and hence a broad trend for loop flow residuals cannot be formalized here.

Scatter plots of studentized residuals

Scatter plots of the residuals, $\left| H\hat{x} - z \right|$, versus the fitted values, $H\hat{x}$, help investigate model inadequacies such as inconsistencies in variance and the inapplicability of a linear model. These inferences can be drawn by comparing the patterns of the scatter plot with approximate shapes depicting particular inadequacies, e.g., a departure from a linear relation is indicated by a curvilinear trend in the scatter plot [57]. Figure 35 and Figure 36 show the scatter plots for the estimates of loop flows using individual schedule of WGENCO at bus 5 and the aggregate schedule, respectively. The scatter plots shown are one particular case in the MC iterations, chosen as representative. In this case, the pattern of scatter plots for residuals of estimated loop flows obtained from individual and aggregate schedules indicate no fixed pattern except for a few potential outliers. For the individual schedule case, the outlier identified from ordinary residuals is found to be in acceptable range when plotted as a studentized residual. The nature of scatter plot in Figure 36 shows the same set of studentized residuals versus respective fitted values with point accurate forecast, lower and upper bounds estimates of WGENCO. All the three subplots have no discernible patterns. Potential outliers are recognized and highlighted. Similar potential outliers may be observed in the residual scatter plots estimated loop flows

obtained by using from individual schedules of other WGENCOs. However, detailed analysis of a potential outlier and its cause will be needed for such a classification.
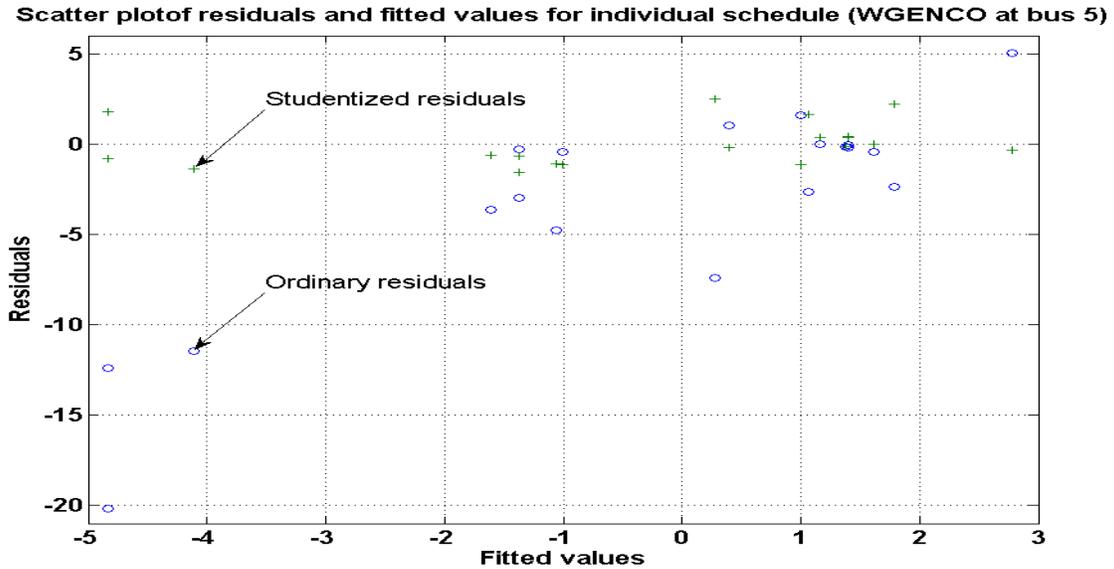

Figure 35 Scatter plot of ordinary and studentized residuals of loop flow estimates obtained from the individual schedSule of WGENCO at bus 5.
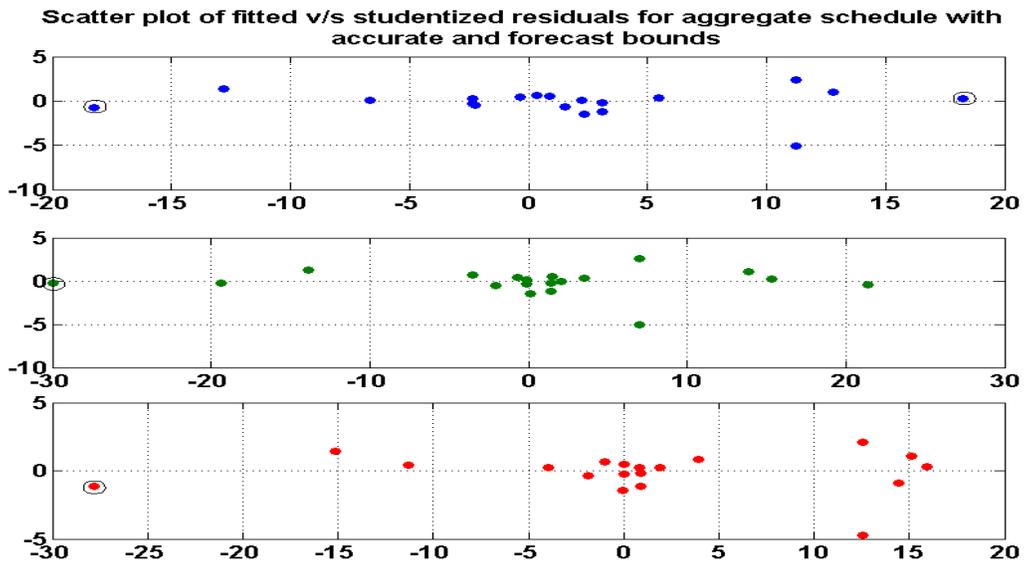

Figure 36 Scatter plot of studentized residuals of loop flow estimates obtained from the aggregate schedule with accurate point estimate (blue), lower bound (green), and upper bound (red) of WGENCO outputs.

Nature of distributions of estimated loop flow

The Kolmogorov-Smirnov (KS) test is used to compare the distributions of estimated loop flows obtained under the three WGENCO output scenarios of accurate point estimate, lower and upper bound of estimates. For simplicity, only the distributions of estimated loop flows obtained from the aggregate schedule are analyzed. However, the approach is extendible to loop flow distributions obtained from individual transactions/schedules. The KS test is performed to infer the inequality among estimated loop flows such that the estimated loop flows obtained from upper and lower bounds of estimates (test distributions) are each compared to the loop flows obtained from accurate point estimates of WGENCO output (reference distribution). The null hypothesis is that the CDF of the reference distribution of the estimated loop flows, $CDF_{(reference)}$, is smaller than that of the test distribution, $CDF_{(test)}$. Consequently, the alternative hypothesis is that the CDF of the test distribution is smaller than that of the reference distribution. A flag value of '1' indicates that KS test rejects the null hypothesis with a 5% significance level; and a flag value of '0' indicates the KS test fails to reject the null hypothesis. as a flag value of '1' or '0'. For a potential value of loop flow, the KS test checks for the outcome of $CDF_{(test)} < CDF_{(reference)}$ and provides the result by comparing the test statistic with the assumed level of significance, in this case, 5%. The impact of forecasting error leads to a horizontal displacement of distribution of estimated loop flows. A flag of '1' implies that the proportion of $CDF_{(test)}$ less than or equal to given loop flow value is less than the proportion of $CDF_{(reference)}$ that is less than or equal to the same given value of loop flow. Analogous interpretation of flag of '0' can be drawn. As indicated by the flags, except for the estimated loop flows 2 and 4 corresponding to both the lower and upper bound estimates of the WGENCO outputs, the null hypothesis is rejected. In other words, only CDFs of estimated loop flows 2 and 4 obtained from lower and upper bounds

of WGENCO output are larger than the respective CDFs of estimated loop flows obtained from accurate point estimates of WGENCO output. The $CDF_{(test)}$ may be displaced in either left or right with respect to the $CDF_{(reference)}$. The test statistic can be used for the quantification of the displacement of the $CDF_{(test)}$. The distributions of forecasting error of the WGENCO output are conservative and possess larger variances; the displacement of $CDF_{(test)}$ is non-trivial for the test case presented. A better modeling of forecasting error in the WGENCO output will lead to smaller displacement of the test distributions and may provide tighter bounds on expected interval estimates of the loop flows.

Table 18 displays the statistics of the estimated loop flow distributions and output of KS tests that compares test distribution with reference distribution as a flag value of '1' or '0'. For a potential value of loop flow, the KS test checks for the outcome of $CDF_{(test)} < CDF_{(reference)}$ and provides the result by comparing the test statistic with the assumed level of significance, in this case, 5%. The impact of forecasting error leads to a horizontal displacement of distribution of estimated loop flows. A flag of '1' implies that the proportion of $CDF_{(test)}$ less than or equal to given loop flow value is less than the proportion of $CDF_{(reference)}$ that is less than or equal to the same given value of loop flow. Analogous interpretation of flag of '0' can be drawn. As indicated by the flags, except for the estimated loop flows 2 and 4 corresponding to both the lower and upper bound estimates of the WGENCO outputs, the null hypothesis is rejected. In other words, only CDFs of estimated loop flows 2 and 4 obtained from lower and upper bounds of WGENCO output are larger than the respective CDFs of estimated loop flows obtained from accurate point estimates of WGENCO output. The $CDF_{(test)}$ may be displaced in either left or right with respect to the $CDF_{(reference)}$. The test statistic can be used for the quantification of the displacement of the $CDF_{(test)}$. The distributions of forecasting error of the WGENCO output are

conservative and possess larger variances; the displacement of $CDF_{(test)}$ is non-trivial for the test case presented. A better modeling of forecasting error in the WGENCO output will lead to smaller displacement of the test distributions and may provide tighter bounds on expected interval estimates of the loop flows.

Table 18: Statistics of Estimated Loop Flows for Aggregate Schedules

Loop	Accurate point estimate		Lower bound estimate			Upper bound estimate		
No.	μ	σ	μ	σ	Flag	μ	σ	Flag
1	3.40	29.71	-1.34	31.65	1	4.61	103.1	1
2	-12.00	6.39	-22.30	6.90	0	-16.22	31.42	0
3	3.84	54.54	7.50	59.60	1	5.38	28.26	1
4	-12.30	6.38	-26.20	7.06	0	-20.2	33.60	0
5	-1.43	45.90	5.61	49.10	1	4.25	20.02	1
6	7.04	60.06	9.60	62.43	1	7.85	26.09	1
7	16.60	29.10	20.30	31.42	1	19.82	8.48	1
8	1.88	45.72	6.83	45.57	1	5.24	20.12	1

Normal distribution of forecasting error associated with WGENCO output and wind variability both impart stochasticity to the actual line flows and hence is observed in the loop flow estimates. Determining the distribution of the estimated loop flows for each case is an exacting exercise. A chi-square test is performed to check the goodness-of-fit of the estimated loop flows corresponding to the aggregate schedule with normal distribution of statistics listed in as a flag value of '1' or '0'. For a potential value of loop flow, the KS test checks for the outcome of $CDF_{(test)} < CDF_{(reference)}$ and provides the result by comparing the test statistic with the

117

assumed level of significance, in this case, 5%. The impact of forecasting error leads to a horizontal displacement of distribution of estimated loop flows. A flag of '1' implies that the proportion of $CDF_{(test)}$ less than or equal to given loop flow value is less than the proportion of $CDF_{(reference)}$ that is less than or equal to the same given value of loop flow. Analogous interpretation of flag of '0' can be drawn. As indicated by the flags, except for the estimated loop flows 2 and 4 corresponding to both the lower and upper bound estimates of the WGENCO outputs, the null hypothesis is rejected. In other words, only CDFs of estimated loop flows 2 and 4 obtained from lower and upper bounds of WGENCO output are larger than the respective CDFs of estimated loop flows obtained from accurate point estimates of WGENCO output. The $CDF_{(test)}$ may be displaced in either left or right with respect to the $CDF_{(reference)}$. The test statistic can be used for the quantification of the displacement of the $CDF_{(test)}$. The distributions of forecasting error of the WGENCO output are conservative and possess larger variances; the displacement of $CDF_{(test)}$ is non-trivial for the test case presented. A better modeling of forecasting error in the WGENCO output will lead to smaller displacement of the test distributions and may provide tighter bounds on expected interval estimates of the loop flows.

Table 18The null hypothesis is that the estimated loop flows possess a normal distribution with known statistics. The alternative hypothesis is that the estimated loop flows are not normally distributed. Table 19 displays the results of the test with flag '1' indicating rejection of the null hypothesis and a flag '0' indicating a failure to reject the null hypothesis. The level of significance for rejection of the null hypothesis by the chi-square test is 5%. Loop flow estimates of the representative case from the aggregate scheduled are also listed in the Table 19appropriately.

Table 19: Outcomes of Chi-square Test for Goodness-Of-Fit of Estimated Loop flows

Loop No.	Accurate point estimate		Lower bound estimate		Upper bound estimate	
	Flag	$\hat{x}_{rr}$ (MW)	Flag	$\hat{x}_{rr}$ (MW)	Flag	$\hat{x}_{rr}$ (MW)
1	1	9.10	1	10.31	1	10.85
2	1	-8.83	1	-16.10	1	-13.03
3	1	-4.39	1	1.25	1	-3.59
4	1	-9.17	1	-19.62	1	-16.95
5	1	-7.61	1	1.81	1	-2.52
6	1	-2.16	1	3.30	1	-1.70
7	1	10.64	1	17.14	1	13.40
8	1	-5.28	1	1.91	1	-2.57

For all the 24 distributions of the estimated loop flows, the flags indicate departure from normality. Departure from normality is also expected in case of distributions of the estimated loop flows obtained from individual schedules. A simple visual inspection reveals that such distributions lack symmetry and might be bimodal. An in-depth investigation to determine the accurate standard distribution type may be required for further inferences.

5.6 Concluding remarks

The impacts of forecasting error distribution (assumed normal) on the estimates of minor loop flows are presented with the help of a case study on the IEEE 14 bus test system. Assuming a positive error correlation, the interval estimates of forecasted wind power values are obtained using a likely interval of the mean absolute error. A detailed analysis of the applicability of linear model to estimation of loop flows under heavy wind energy penetration is performed.

Adherence of studentized residuals to normality, validity of linear relation, and light tailed distributions are the major observations. Existence of mild outliers is suspected. Outliers when re-plotted with studentized residuals exhibit acceptable deviation. However, an outlier in any case will require a systematic analysis and detailed justification. Finally, KS test and chi-square goodness-of-fit tests were performed to interpret the nature of distributions of estimated loop flows. Lack of normality in distribution of estimated loop flows and a generic decrease in CDF of estimated loop flows as a result of forecasting error was observed in the case study. Generic formal models and interpretations are foreseen as future tasks. The outcome of this analysis may be employed in allocating USF responsibility to WGENCOs as an interval estimate.

CHAPTER 6

CONCLUDING REMARKS AND FUTURE WORK

6.1   Concluding remarks

Variable USFs in transmission networks are modeled using a mathematical artifact *minor loop flows* in the proposed framework. A linear estimator from the literature is adopted that relates USF to minor loop flow estimates via an incidence matrix. The primary cause of USFs is the fundamental inconsistency between market-expected flows (computed based on superposition theorem) and actual flows (estimated using non-linear power flow equations). The variability in USF is attributed to the significant levels of wind energy penetration at the transmission level and forecasting errors associated with it. Multiple techniques are adopted in order to automate the computation and analysis of loop flows.

The proposed framework is presented for identifying the implications of variable USF from both power systems-planning and market operations perspective. For the analysis of USF in practical transmission networks, human decision-making in the form of visually interpreting the incidence matrix is not feasible. Detection of accurate loops (i.e., system observation) and determination of transmission line layouts (i.e., directional information) are vital steps for the synthesis of incidence matrix. This dissertation proposes a generic LDA with an objective of detecting closed trails in a transmission network. The search algorithm is constrained by allowing each transmission line in the first active network to be traversed once each in opposite directions. The output of LDA is a set of closed trails (loops) in the network and is an essential observation for modeling the minor loop flow*s*. The algorithm is flexible to synthesize and accommodate changes in the topology of the transmission network (i.e., line outages and transmission expansions). The algorithmic approach has distinct advantages over the

121

contemporary graph theory approaches of modeling loop flows such as requiring no visual aids; accounting for all the transmission lines; and, selecting loops without the need of power flow information.

The incidence matrix which maps minor loops to transmission lines requires line layout information. A geographical information systems (GIS) application is proposed to extract the line layout information in the form of *azimuth* of transmission lines. GIS techniques assist in the automated synthesis of the incidence matrix hence enabling the framework applicable to practical transmission networks. Additionally, as information on GIS coordinates of buses is a standard component of power systems databases, no specialized input is needed.

Monte Carlo (MC) simulations are used to solve the probabilistic load flow analysis with the wind power as the stochastic variable. MC simulation is an established technique to achieve maximum accuracy in such probabilistic experiments. The solution of linear estimator provides the values of minor loop flow estimates. Ordinary least squares, analytic ridge regression, and robust regression (M-estimators) are used to estimate loop flows. An application of building the regression model − model reduction (MR) − is demonstrated. MR is based on constraints accounting both statistical significance of regressors and power system operations. The quality of the minor loop flow estimates is also ensured acceptable by using suitable automated techniques.

Power output of wind power plants is subject to forecasting errors dependent on the horizon of prediction. A day-ahead market simulation is analyzed using a normal distribution modeling forecasting error in wind power output. Prediction intervals of wind power output are proposed to be more effective than point forecasts and their impacts on variable USF are

investigated. Statistical inferences and linear model adequacy to accommodating variable USF are also documented.

6.2   Scope of future work

An immediate avenue of future work –deals with the iterations in LDA when changes in decision are required to resolve inconsistent solutions. As explained in chapter 2, LDA encounters non-deterministic cases and hence takes encumbers the user to take decisions. There are three instances when the LDA takes decisions namely: solution of vertex with degree two and four distinct unidirectional edges; determination of the starting vertex; and selection of the outward direction for wave search. These three decisions have an influence on the consequent steps and hence the final solution. Any decisions along with their corresponding alternatives need to be stored. The optimum way to store such information is using bounded pointers.

First instance of decision-making is shown in Figure 8 on page 44, and has two equally correct solutions – forming either composite edges [4-1-3] and [2-1-5] or alternatively [4-1-5] and [2-1-3]. After selecting a solution out of the two choices, the active network information, and the other alternative should be stored. This information may be recalled and changed in case the LDA encounters an invalid solution due to this decision. The second instance of decision-making is the selection of the starting vertex from a set of vertices obtained from successive summation of weights. It is commonly observed that successive summation process of weights of vertices to determine a minimal locally connected vertex leads to multiple potential starting vertices. In such instances, an arbitrary choice of the starting vertex usually suffices the purpose. However, it is recommended to store the selected starting vertex, active network configuration, and other potential starting vertices, if any. The last instance of decision-making is the choice of outward direction for the wave search chosen from the starting vertex. For example, in Figure 9 on page

45 the outward direction 6-1 is arbitrarily chosen for solving the vertex and detecting a loop. The potential outward directions 6-5 and 6-7 are topologically correct choices and lead to distinct loops. In fact, the outward direction 6-5, if selected, leads to the loop with vertices {6-5-8-7-6}. This solution has lesser number of vertices as compared to the loop obtained from the arbitrary outward direction of 6-1, and hence adheres to the objective of forming loops with minimum length. Given the sparse nature of transmission networks, a localized optimization of outward direction selection may be a feasible approach to detect the loop with lowest number of vertices. This concludes a brief account of possible refinement in the LDA to improve the performance and obtain better results.

An automated framework to estimate loop flows to model variable USF in an electric grid is proposed in this dissertation. Variability associated with USF is attributed to significant wind power penetration in the network. However, the framework is applicable to utility level solar power plants without major modifications. The outcome of the proposed research is potentially suitable for both market and network level accommodation for variable USF. Both the applications are briefly explained in this subsection.

A market level accommodation of variable USF from a WGENCO perspective can be developed on the basis of the loop flow estimates. This accommodation is analogous to the GENCO-oriented USF accommodation in [2]. A contribution factor (CF) to allocate USF responsibility to GENCOs is shown in equation (9) [2]:

$$CF_m = \frac{\left\| \hat{x}_{agg} - \left[\sum_{i=1:M}^{i \neq m} \hat{x}_i \right] \right\|_P}{\left\| \hat{x}_{agg} \right\|_P} \times \frac{gen(m)}{sum(gen)} \qquad (9)$$

124

where, CF_m is the contribution factor for m[th] GENCO, $\hat{x}_{agg}$ are the loop flow estimates corresponding to the aggregate schedule, $\hat{x}_i$ are the loop flow estimates for individual schedules of generators, P is the norm, and $gen(m)$ is the power output of the mth GENCO, $sum(gen)$. The above contribution factor accounts for the USF responsibility of conventional generators with fixed (deterministic) power output. The contribution of a WGENCO can be computed by a suitable modification of (6.2). Contribution factor for a WGENCO needs to account for the contract period, energy metering mechanism, and the variable USF. For a WGENCO participating in a market with a resolution of 10 minutes the contribution factor can be computed as shown in equation (10):

$$CF_m = \sum_{j=1}^{10} \frac{\left\| \hat{x}_{agg} - \left[\sum_{i=1:M}^{i \neq m} \hat{x}_{i,j} \right] \right\|_P}{\left\| \hat{x}_{agg} \right\|_P} \times \frac{gen(m)}{sum(gen)} \tag{10}$$

where, $\hat{x}_{i,j}$ is the estimated loop flow for each minute depending on the actual power output of the WGENCO, i is the count of generators, and j is the time resolution of wind power output measurement. The market expectations in this case should be computed using the approved e-tags. The contribution factor for WGENCO may be used in conjunction with any transmission pricing paradigm, similar to that of the GENCO contribution factor proposed in [2]. This concludes the market accommodation of variable USF in a SMD with known e-tags.

Network level accommodation of variable USF includes use of QCD at strategically critical buses in the grid. The use of QCD can assist in controlling the power flow on critical lines and hence counter congestion and variable USF [14]. The location and operation of QCDs can be optimized by using the research proposed in this dissertation. Location of a QCD can be recognized in a systems-planning approach, primarily by observing historical values of loop flow estimates obtained using the proposed research. An alternative approach is to select the location

125

of a bus associated with a critical line (see chapter 4). Ideally a QCD should reduce the USF on a transmission line to a zero. However, practically QCDs reduce USF on critical transmission lines to an acceptable level and cause USF on other lines. The z vector (and hence estimated loop flows) in the linear estimator reflects the changes on USF due to the use of QCDs with the rest of the computation unchanged. Thus, location and operational strategies of QCDs can be optimized using the research proposed in this dissertation.

The computation of variable loop flows may find application in transmission planning, congestion analysis, schedule optimization, and resource management. Importance of analysis of USF (a critical seams issue) in power systems expansion and planning is gaining wide recognition. Hence, a holistic and automated mathematical framework to estimate variable loop flows is proposed in this dissertation which may have multiple applications.

REFERENCES

[1] US DOE, "20% wind energy by 2030: Increasing wind energy's contribution to US electricity supply," 2008. [Online]. Available: www.nrel.gov/docs/fy08osti/41869.pdf. [Accessed: 10-Aug-2011].

[2] S. Suryanarayanan, "Accommodation of loop flows in competitive electric power systems," PhD Dissertation, Dept. of ECE, Arizona State University, Tempe, AZ, 2004.

[3] D. S. Kirschen and G. Strbac, *Fundamentals of power system economics*. Chichester, West Sussex, England: John Wiley & Sons, 2004.

[4] Anonymous, "Potential use of Firm Contingent energy product code on e-Tags for variable generation For transparency under WECC product codes," 29-Apr-2010. [Online]. Available: http://www.wspp.org/filestorage/firm_contingent_discussion_paper_042610.pdf. [Accessed: 30-Jul-2013].

[5] "General agreement on parallel paths experiment." [Online]. Available: http://www.nerc.com/docs/docs/pubs/GAPP_exper_report_1999.pdf. [Accessed: 02-Dec-2012].

[6] A. Clamp, "The lake Erie loop flow mystery," Jan-2010. [Online]. Available: http://appanet.cms-plus.com/files/PDFs/LakeErieLoopFlowMysteryPublicPowerJanFeb2010.pdf. [Accessed: 10-Dec-2012].

[7] W. Hogan, "Interregional Coordination of Electricity Markets," presented at the Technical Conference on Interregional Coordination, Washington DC.

[8] R. Schuler, et al., "Electricity Market Structures to Reduce Seams and Enhance Investment," Feb-2010. [Online]. Available: http://certs.lbl.gov/pdf/schuler-M9.pdf. [Accessed: 12-Aug-2013].

[9] K. Spees and J. Pfeifenberger, "Seams Inefficiencies - Problems and Solutions at Energy Market Borders," 17-Jul-2012. [Online]. Available: http://www.brattle.com/_documents/UploadLibrary/Upload1057.pdf. [Accessed: 07-Aug-2013].

[10] "MISO Strategive Initiatives - Seams Issues." [Online]. Available: https://www.misoenergy.org/WhatWeDo/StrategicInitiatives/Pages/Seams.aspx. [Accessed: 02-Oct-2013].

[11] M. Adibi, "Inquiry about the 14 Bus IEEE Test System," 05-Mar-2012.

[12] "pg_tca14bus." [Online]. Available: http://www.ee.washington.edu/research/pstca/pf14/pg_tca14bus.htm. [Accessed: 19-Mar-2013].

[13] M. Albaijat, K. Aflaki, and B. Mukherjee, "Congestion management in WECC grid," in *Proc. 2012 IEEE PES, Innovative Smart Grid Technologies (ISGT)*, pp. 1 –8.

[14] P. Guha Thakurta, D. Van Hertem, and R. Belmans, "An approach for managing switchings of controllable devices in the Benelux to integrate more renewable sources," in *Proc. 2011 IEEE, Trondheim PowerTech*, pp. 1–7.

[15] J. A. Kavicky and S. M. Shahidehpour, "Parallel path aspects of transmission modeling," *IEEE Trans. Power Systems*, vol. 11, no. 3, pp. 1180–1190, Aug. 1996.

[16] Anonymous, "General agreement on parallel paths experiment," 31-Aug-1999. [Online]. Available: www.nerc.com/docs/docs/pubs/GAPP_exper_report_1999.pdf. [Accessed: 13-Jul-2011].

[17] C. M. Davis, J. D. Weber, and K. Johnson, "Circulating MW and Mvar flows in large systems," *North American Power Symposium (NAPS)*, pp. 1–6, Jun. 2009.

[18] S. Cvijic, M. Cvetkovic, and M. Ilic, "A graph-theoretic approach to modeling network effects of phase shifters on active power loop flows," *North American Power Symposium (NAPS)*, pp. 1–6, Sep. 2012.

[19] D. Das and D. Divan, "Individual generator contributions towards loads and line flows in networks with loop flows," *North American Power Symposium (NAPS)*, pp. 1–6, 2009.

[20] S. Suryanarayanan, "Techniques for accommodating unscheduled flows in electricity networks and markets," *IEEE PES General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pp. 1–6, 2008.

[21] S. Suryanarayanan and G. T. Heydt, "Modification to Contribution Factor Formula for Unscheduled Flows," *IEEE Trans. Power Systems*, vol. 23, no. 2, pp. 809 –810, May 2008.

[22] D. Hayward, J. Miller, B. Balmant, K. Morris, J. Malinowski, B. Pasternack, L. Eilts, "Operating problems with parallel flows," *IEEE Trans. Power Systems*, vol. 6, no. 3, pp. 1024 –1034, Aug. 1991.

[23] Anonymous, "Unscheduled Flow Administrative Subcommittee," 13-Dec-2002. [Online]. Available:
http://www.wecc.biz/library/Documentation%20Categorization%20Files/Retire%20and%20Archive/Administrative%20Practice%20008%20-%20Use%20of%20Phase%20Shifting%20Transformers%20to%20Provide%20Obligated%20Path%20Relief.pdf. [Accessed: 03-Mar-2013].

[24] W. J. Lyman, "Controlling power flow with phase shifting equipment," *Trans. American Institute of Electrical Engineers*, vol. 49, no. 3, pp. 825–829, Jul. 1930.

[25] Y. R. Sood, N. P. Padhy, and H. O. Gupta, "Wheeling of power under deregulated environment of power system-a bibliographical survey," *IEEE Trans. Power Systems*, vol. 17, no. 3, pp. 870–878, Aug. 2002.

[26] "Tagging Essentials For Etag 1.7" [Online]. Available: http://reg.tsin.com/Tagging/e-tag/Tagging%20Essentials.pdf. [Accessed: 10-Aug-2011].

[27] J. Kritikson, "California Electricity Market Primer," Feb-2000. [Online]. Available: http://www.ucei.berkeley.edu/Restructuring%20Archive/Cal_Elec_Primer.pdf. [Accessed: 10-Jul-2013].

[28] P. L. Joskow, "Markets for power in the United States: An interim assessment," 2005. [Online]. Available: http://dspace.mit.edu/handle/1721.1/45042. [Accessed: 01-Oct-2013].

[29] S. Suryanarayanan, R. G. Farmer, G. T. Heydt, and S. Chakka, "Estimation of unscheduled flows and contribution factors based on L_p norms," *IEEE Trans. Power Systems*, vol. 19, no. 2, pp. 1245–1246, May 2004.

[30] S. Suryanarayanan and G. T. Heydt, "A linear static Kalman filter application for the accommodation of unscheduled flows," *IEEE PES Power Systems Conf. and Exposition*, pp. 179–184 vol.1, 2004.

[31] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE Trans. Systems Science and Cybernetics*, vol. 4, no. 2, pp. 100–107, Jul. 1968.

[32] D. Maier, "An Efficient Method for Storing Ancestor Information in Trees," *SIAM Journal on Computing*, vol. 8, no. 4, pp. 599–618, 1979.

[33] A. Tucker, *Applied combinatorics*. Hoboken, NJ: Wiley, 2007.

[34] M. Mohanpurkar, D. J. Zimmerle, and S. Suryanarayanan, "An algorithmic approach to tracing closed loops in power systems network," *IEEE Trans. Power Systems, Under review*, p. 8.

[35] M. Mohanpurkar, H. Valdiviezo, and S. Suryanarayanan, "An application of geographical information systems technique in the estimation of loop flows in wide area transmission networks with high wind penetration," in *45th Proc. of Frontiers of Power Conf.*, Oklahoma, 2012, pp. VI – 1 – VI – 8.

[36] *Wind power in power systems*. Chichester, West Sussex: Wiley, 2012.

[37] Anonymous, "Representing the global wind energy industry," *Global Wind Energy Council*. [Online]. Available: http://www.gwec.net/global-figures/graphs/. [Accessed: 15-Mar-2013].

[38] R. Wiser and M. Bolinger, "2011 Wind Technologies Market Report." [Online]. Available: http://www1.eere.energy.gov/wind/pdfs/2011_wind_technologies_market_report.pdf. [Accessed: 15-Mar-2013].

[39] A. Botterud, A. Zhou, J. Wang, R. Bessa, H. Keko, J. Mendes, J. Sumaili, and V. Miranda, "Use of Wind Power Forecasting in Operational Decisions," *Argonne National Laboratory*, 2011. [Online]. Available: http://www.dis.anl.gov/pubs/71389.pdf. [Accessed: 20-Nov-2012].

[40] R. Doherty and M. O'Malley, "A new approach to quantify reserve demand in systems with significant installed wind capacity," *IEEE Trans. Power Systems*, vol. 20, no. 2, pp. 587–595, May 2005.

[41] B. Hodge and M. Milligan, "Wind power forecasting error distributions over multiple timescales," in *IEEE PES General Meeting*, 2011, pp. 1–8.

[42] H. Bludszuweit, J. A. Dominguez-Navarro, and A. Llombart, "Statistical Analysis of Wind Power Forecast Error," *IEEE Trans. Power Systems*, vol. 23, no. 3, pp. 983–991, Aug. 2008.

[43] A. Khosravi, S. Nahavandi, D. Creighton, and R. Naghavizadeh, "Uncertainty quantification for wind farm power generation," in *The 2012 International Joint Conf. on Neural Networks (IJCNN)*, June, pp. 1–6.

[44] M. Mohanpurkar and S. Suryanarayanan, "A case study on the effects of predicted wind farm power outputs on unscheduled flows in transmission networks," in *5th Annual Green Technologies Conf.*, Colorado, 2013, pp. 277–284.

[45] J. D. Glover, *Power system analysis and design*. Stamford, CT: Cengage Learning, 2012.

[46] P. Sauer, "A generalized stochastic power flow algorithm," Dept. of ECE, Purdue, West Lafayette, IN.

[47] R. N. Allan, B. Borkowska, and C. H. Grigg, "Probabilistic analysis of power flows," *Proc. of the Institution of Electrical Engineers*, vol. 121, no. 12, pp. 1551–1556, Dec. 1974.

[48] Chun-Lien Su, "Probabilistic load-flow computation using point estimate method," *IEEE Trans. Power Systems*, vol. 20, no. 4, pp. 1843–1851, Nov. 2005.

[49] P. Jorgensen, J. S. Christensen, and J. O. Tande, "Probabilistic load flow calculation using Monte Carlo techniques for distribution network with wind turbines," in *Proc. of 8th International Conf. on Harmonics and Quality of Power*, 14, vol. 2, pp. 1146–1151 vol.2.

[50] R. N. Allan and M. R. G. Al-Shakarchi, "Probabilistic a.c. load flow," *Proc. of the Institution of Electrical Engineers*, vol. 123, no. 6, pp. 531–536, Jun. 1976.

[51] R. N. Allan and M. R. G. Al-Shakarchi, "Probabilistic techniques in a.c. load-flow analysis," *Proc. of the Institution of Electrical Engineers*, vol. 124, no. 2, pp. 154–160, Feb. 1977.

[52] F. Alvarado, Y. Hu, and R. Adapa, "Uncertainty in power system modeling and computation," *IEEE International Conf. on Systems, Man and Cybernetics*, pp. 754–760 vol.1, 1992.

[53] J. Schwippe, O. Krause, and C. Rehtanz, "Extension of a probabilistic load flow calculation based on an enhanced convolution technique," *IEEE PES/IAS Conf. on Sustainable Alternative Energy (SAE)*, pp. 1–6, 2009.

[54] G. Golub, *Matrix computations*. Baltimore: Johns Hopkins University Press, 2013.

[55] E. Kyriakides, S. Suryanarayanan, and G. T. Heydt, "State estimation in power engineering using the Huber robust regression technique," *IEEE Trans. Power Systems*, vol. 20, no. 2, pp. 1183–1184, May 2005.

[56] M. Mohanpurkar and S. Suryanarayanan, "Accommodating unscheduled flows in electric grids using analytical ridge regression," *IEEE Trans. on Power Systems*, vol. 28, no. 3, pp. 3507–3508, Aug. 2013.

[57] *Applied linear statistical models*. Boston: McGraw-Hill Irwin, 2005.

[58] *Ibid.* [57], pp. 437-449.

[59] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics-Theory and Methods*, vol. 6, no. 9, pp. 813–827, 1977.

[60] D. C. Montgomery, *Introduction to linear regression analysis*. Hoboken, N.J.: Wiley-Interscience, 2006.

[61] F. J. Massey Jr, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[62] "Chi-square goodness-of-fit test - MATLAB chi2gof." [Online]. Available: http://www.mathworks.com/help/stats/chi2gof.html. [Accessed: 26-Mar-2013].

[63] M. Mohanpurkar and S. Suryanarayanan, "Regression Modeling for Accommodating Unscheduled Flows in Electric Grids," *IEEE Trans. Power Systems (Under review)*.

[64] "Simulator SimAuto, PowerWorld." [Online]. Available: http://www.powerworld.com/products/simulator/add-ons-2/simauto. [Accessed: 21-Mar-2013].

[65] K. Iba, "Identification of transmission line user and congestion management by loop flow controllers," *CIGRE/IEEE PES, International Symposium*, pp. 307–314, Oct. 2005.

[66] S. Heier, *Grid integration of wind energy conversion systems*. Chichester, England ;: Wiley, 2006.

[67] Andre Dozier, Joseph Liberatore, and Matthew Veghte, "Unscheduled flow within deregulated electricity markets: Bridging the gap between the western electric power industry and academia," in *IEEE PES General Meeting*, San Diego, CA, 2012.

[68] P. R. Gribik, D. Shirmohammadi, J. S. Graves, and J. G. Kritikson, "Transmission rights and transmission expansions," *IEEE Trans. Power Systems*, vol. 20, no. 4, pp. 1728–1737, Nov. 2005.

[69] S. T. Mak and N. Farah, "Synchronizing SCADA and smart meters operation for advanced smart distribution grid applications," in *IEEE PES Innovative Smart Grid Technologies (ISGT)*, 2012.

[70] T. Kristiansen, "Provision of financial transmission rights," *in Proc. of IEEE Power Tech Conf. 2003*, vol. 4, pp. 1–10.

[71] C. Y. Choo, N.-K. C. Nair, and B. Chakrabarti, "Impacts of Loop Flow on Electricity Market Design," *International Conf. on Power System Technology, (PowerCon)*, pp. 1–8, 2006.

[72] R. Baldick and E. Kahn, "Contract paths, phase-shifters, and efficient electricity trade," *Power Systems, IEEE Transactions on*, vol. 12, no. 2, pp. 749–755, May 1997.

[73] G. Granelli, M. Montagna, F. Zanellini, P. Bresesti, and R. Vailati, "A genetic algorithm-based procedure to optimize system topology against parallel flows," *IEEE Trans. Power Systems*, vol. 21, no. 1, pp. 333 – 340, Feb. 2006.

[74] H. Stark and J. W. Woods, *Probability, random processes, and estimation theory for engineers*. Prentice-Hall Englewood Cliffs (NJ), 1986.

[75] P. Jorgensen, J. S. Christensen, and J. O. Tande, "Probabilistic load flow calculation using Monte Carlo techniques for distribution network with wind turbines," in *Proc. of 8th International Conf. on Harmonics and Quality of Power*, 1998, vol. 2, pp. 1146–1151.

[76] A. E. Hoerl, R. W. Kannard, and K. F. Baldwin, "Ridge regression: some simulations," *Communications in Statistics-Theory and Methods*, vol. 4, no. 2, pp. 105–123, 1975.

[77] S. Even, "An algorithm for determining whether the connectivity of a graph is at least k," *SIAM Journal on Computing*, vol. 4, no. 3, pp. 393–396, 1975.

[78] J. Hopcroft and R. Tarjan, "Dividing a graph into triconnected components," *SIAM Journal on Computing*, vol. 2, no. 3, pp. 135–158, 1973.

[79] E. Cotilla-Sanchez, P. D. H. Hines, C. Barrows, and S. Blumsack, "Comparing the topological and electrical structure of the North American electric power infrastructure," *IEEE Systems Journal*, vol. 6, no. 4, pp. 616–626, Dec. 2012.

[80] T. Asano, "An application of duality to edge-deletion problems," *SIAM Journal on Computing*, vol. 16, no. 2, pp. 312–331, 1987.

[81] T. Zhong, "Graph theory based expert system to form de-icing route in Changsha power grid," *International Journal of Electrical Power & Energy Systems*, vol. 43, no. 1, pp. 1318–1321, Dec. 2012.

[82] "FERC: Major Orders & Regulations." [Online]. Available: http://www.ferc.gov/legal/maj-ord-reg.asp. [Accessed: 19-Sep-2012].

[83] D. S. Kirschen and G. Strbac, *Fundamentals of Power System Economics*, 1st ed. Wiley, 2004.

[84] Anonymous, "WSCC Unscheduled Flow Mitigation Plan," 30-Jul-2001. [Online]. Available: http://www.wecc.biz/committees/standingcommittees/oc/ufas/shared%20documents/ufas%20mitigation%20plan.pdf. [Accessed: 10-Oct-2011].

[85] R. Baldick, J. Bushnell, S. Tierney, T. Winter, and A. Brown, "A National Perspective on Allocating the Costs of New Transmission Investment: Practice and Principles," *Working Group for Investment in Reliable and Economic Electric Systems (WIRES)*, 2007.

[86] V. S. C. Lim, J. D. F. McDonald, and T. K. Saha, "Application of loop frame of reference to power flow tracing and loss allocation," in *The 7th International Power Engineering Conf. (IPEC)*, 2005, pp. 1013 –1018 vol. 2.

[87] S. Suryanarayanan, G. T. Heydt, R. G. Farmer, and S. Chakka, "An estimation technique to contribution factors for loop flows in an interconnected power system," *Electric power components and systems*, vol. 19, no. 2, pp. 813–826, 2004.

[88] R. Diestel, *Graph theory*. New York: Springer, 2000.

[89] M. Yannakakis, "Edge-deletion problems," *SIAM Journal on Computing*, vol. 10, no. 2, pp. 297–309, 1981.

[90] T. Kong, H. Cheng, Z. Hu, and L. Yao, "Multiobjective planning of open-loop MV distribution networks using ComGIS network analysis and MOGA," *Electric Power Systems Research*, vol. 79, no. 2, pp. 390–398, Feb. 2009.

[91] Y. Cai and M.-Y. Chow, "Exploratory analysis of massive data for distribution fault diagnosis in smart grids," in *IEEE PES General Meeting*, 2009, pp. 1–6.

[92] P. A. Parikh and T. D. Nielsen, "Transforming traditional geographic information system to support smart distribution systems," in *IEEE PES Power Systems Conf. and Exposition (PSCE)*, 2009, pp. 1–4.

[93] K. Wu and Z. Zhang, "Research and implementation of smart transmission grids based on WebGIS," in *Second International Conf. on Communication Systems, Networks and Applications (ICCSNA)*, 2010, vol. 1, pp. 302 –307.

[94] G. Bonham-Carter, *Geographic information systems for geoscientists: modelling with GIS*, vol. 13. Pergamon press, 1994.

[95] S. Chakka, S. Suryanarayanan, and G. T. Heydt, "Analysis and estimation of loop (parallel) flows in wide area interconnected power systems," in *Proceedings of 34th North American Power Symposium*, 2002, pp. 176–180.

[96] *Ibid.* [57], pp. 214-248.

[97] *Ibid.* [57]*,* pp.402-405.

[98] *Ibid* [57]*,* pp. 161-165.

[99] W. Wu and C. Wong, "FACTS applications in preventing loop flows in interconnected systems," *IEEE PES General Meeting*, vol. 1, pp. 170–174, Jul. 2003.

[100] K. Porter and J. Rogers, "Status of centralized wind power forecasting in North America," National Renewable Energy Laboratory, Golden, CO, NREL/SR-550-47853, 2010.

[101] A. Fabbri, T. G. S. Roman, J. R. Abbad, and V. H. M. Quezada, "Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market," *IEEE Trans. Power Systems*, vol. 20, no. 3, pp. 1440–1446, Aug. 2005

[102] M. Mohanpurkar, "Framework for investigating impacts of wind energy penetration on unscheduled flows in bulk interconnections," presented at the Center for Research and Education in Wind, University of Colorado, 28-Sep-2012.

APPENDIX

A. Matlab codes for loop detection algorithm (LDA), probabilistic load flow analysis (Monte Carlo simulations), and regression analysis.