USING SENSOR FUSION TECHNIQUES TO IMPROVE VOLTAGE INSTABILITY PREDICTIONS USING LOCAL MEASUREMENTS

A Thesis

by

BENJAMIN DAVID PICONE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Miroslav Begovic
Committee Members,	Chanan Singh
	Gwan Choi
	Behbood Zoghi
Head of Department,	Miroslav Begovic

May 2020

Major Subject: Electrical Engineering

ABSTRACT

Voltage collapse is a risk to power systems that arises when a system is loaded to near its power transfer limits. As grids are operated closer to their limits, the risk of this phenomenon occurring increases. This research is an effort to build upon the concepts of the Voltage Instability Device (VIP) to improve predictions about proximity to voltage collapse. The principle behind the VIP device is to linearize the power system using simplifying assumptions and current and voltage measurements local to a load. An issue encountered by using this method is that the predictions by VIP devices throughout the system tend to vary considerably. A sensor-fusion framework is presented that makes use of multiple inputs from a network of nearby sensors to attempt to improve prediction accuracy. The sensor-fusion framework employed is known as Dempster-Shafer Evidential (DSE) Theory. This theory relies on the assignment of probabilities to represent the support for the evidence presented by each "sensor" (i.e. VIP device) location. In this work, a consensus algorithm is developed using measurements nearby and a centrality algorithm is used to rate how central a VIP location is in the system. The consensus algorithm is shown to consistently improve the overall error in prediction by a moderate amount. The centrality algorithm improves the prediction in some cases, but not in larger systems. Overall the research presents a positive, albeit limited, improvement in VIP accuracy. The DSE framework employed shows promise as a method to combine data, however, improvements on the algorithms for assigning evidence would be needed for truly successful implementation.

CONTRIBUTORS AND FUNDING SOURCES

This work was advised by Prof. Miroslav Begovic, my academic advisor and thesis committee chair. The thesis committed consisted of Prof. Chanan Singh of the ECE department, Prof. Gwan Choi of the ECE department, and Prof. Behbood Zoghi of the ETID department.

Preliminary research that contributed to this work was conducted in conjunction with Aaqib Peerzada of the ECE department. This preliminary research consisted of developing the VIP device model for use in the simulations. The study was partially funded by a fellowship from Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION AND LITERATURE REVIEW

1.1 Overview

The primary operational concern in real-time reliability management of power systems is making sure that sufficient power is available to meet demand without violating the physical constraints of the equipment on the system. These constraints are thermal limits (constraints on the amount of power that can flow through devices), voltage limits (constraints to ensure that connected devices operate at their rated voltages) and contingency constraints (constraints to ensure that if any piece of equipment fails the system will not violate any constraints before re-dispatching can occur). Alongside constraint management, a risk more difficult to identify exists which, if unmitigated, can bring about system collapse. In certain circumstances, this can occur without even violating any serious physical constraints or contingencies. This is known as voltage instability.

The problem of voltage instability stems from the maximum power transfer problem which every engineering student learns about in their introductory circuits class. The problem, simply stated, is that given a system with an internal (or Thévenin) impedance and a load with a varying impedance, the maximum amount of power that can be transferred to the load occurs when the load impedance is equal to the Thévenin impedance of the system. In power systems, when demand increases the effective impedance of the load decreases. Under stable conditions, the demand is met simply by increasing the output of available units that do not violate constraints. However, if the load impedance decreases to a point that it matches the Thévenin impedance of the power system, the demand cannot be met regardless of the availability of generation. Under these conditions, voltage collapse can ensue. While it is a simple phenomenon to describe, it can

be difficult to predict since it does not depend on easily observable metrics, such as the availability of power to meet demand or physical line limits. The difficulty with predicting voltage stability arises since it generally requires a full network representation [1] to model the phenomenon.

In a simple linear circuit, a Thévenin equivalent can be calculated through network reduction: using the properties of the system components derived from Kirchoff's Laws, the individual components with impedance can be modeled as ideal electrical components (resistors, capacitors and inductors) and lumped together to form one equivalent impedance. In power systems, the impedances of the system are represented by an admittance[1] matrix, Y, commonly referred to as the Ybus matrix. The Ybus matrix can be inverted to give the Zbus matrix, Z, of which the diagonal elements, Z_{kk}, give the Thevenin impedance at bus k [2]. However, as researchers in [3-5] point out, Thevenin's theorem does not hold exactly for power systems since they behave as non-linear systems. They overcome this problem by linearizing the system around the operating point by substituting generators' and loads' active and reactive power flows with admittances before inverting the Ybus matrix. Therefore, this technique involves, not only complete knowledge of the system topology, but also complete knowledge of the real-time operating points of every generator and load on the system.

[1] Admittance, y, is the inverse of impedance, z. $y \coloneqq 1/z$. The impedances of a power system are usually represented by the admittance Ybus matrix instead of its inverse, the Zbus matrix. Not only is it simpler to calculate, but also because it is a sparse matrix for large systems, which is easier to store and perform calculations on. Its inverse, the Zbus matrix, is most decidedly not-sparse.

The Thévenin impedance of a black-box system can also be estimated by taking voltage and current measurements at various operating points. If the black-box operates linearly or close to linearly around the observed operating points, then the equivalent internal resistance can be calculated. The Voltage Instability Prediction (VIP) device uses this approach to calculate Thevanin parameters without knowledge of full network representation while still offering robust protection against voltage collapse. Using this linear approximation, the VIP continually estimates the margin of safe operation of the system. In this way the stability of the system can be quickly estimated using only local measurements.

However, as will be seen in this study, the predictions of this operating margin vary depending on the location where the measurements are taken. In order to overcome this problem of disagreement, the goal of this research is to incorporate predictions from nearby buses in order to improve the accuracy of the VIP margin estimate. While the measurements are no longer strictly local, the calculations to improve the accuracy of the predictive ability of the VIP should be simple and fast relative to a centralized voltage stability prediction that requires full system knowledge and computationally intensive simulation.

Sensor fusion is a class of data analysis techniques that use multiple data sources in a similar way that humans' cognitive processes continuously make deductions around observations of the environment [6],[7]. For example, sensory information in the form of sight, smell and touch combine to tell us if food is delicious or rotten. Since information from one sensor might be incomplete or inaccurate, sensor fusion is an attempt to take the advantage of the multiplicity of data to fill in these gaps. Applications are widespread ranging from image-processing [8], medical imaging, target-tracking and identification[9], and robotics [10].

The sensor fusion technique that will be used in this research comes from Dempster-Shafer (D-S) evidential theory. D-S evidential theory allows sensors to contribute detection or classification information based on the extent of their individual knowledge. D-S theory can be interpreted as a generalization of probability theory in which probabilities are assigned to sets instead of mutually exclusive propositions [11].

A key feature is that it allows the assignment of a portion of a sensor's knowledge to uncertainty. Therefore, D-S theory allows decisions about fused data based on an incomplete probabilistic model. Because of its flexibility, D-S theory provides a powerful tool for multi-sensor fusion. For example, a sensor can report with 10% confidence that it is detecting event A, with 50% confidence that it is detecting event B, and with 40% confidence that it has no idea what it is detecting. These confidence measures are referred to as basic probability assignments (BPAs).

One of the key challenges in implementing D-S data fusion techniques lies in successfully identifying a technique for determining BPAs. The resulting fused data is only useful if the BPAs represent confidence in the associated measurements with a degree of accuracy. In this research, various approaches to assigning BPAs were tried and their effectiveness will be judged based on how well they perform on IEEE 39 bus and 118 bus test systems.

1.2 Voltage Stability

When power is insufficient to meet demand, the system frequency drops and generators on automatic generation control use this information as a signal to increase output. The difficulty with voltage stability arises since it requires a full network representation [1].

The problem of voltage instability stems from the maximum power transfer problem. While it is simple phenomenon to describe, it can be difficult to predict since it does not depend on the

availability of power to meet demand. Moreover, the maximum power problem is derived from an infinite bus model, the name itself implying that the power source not limited.

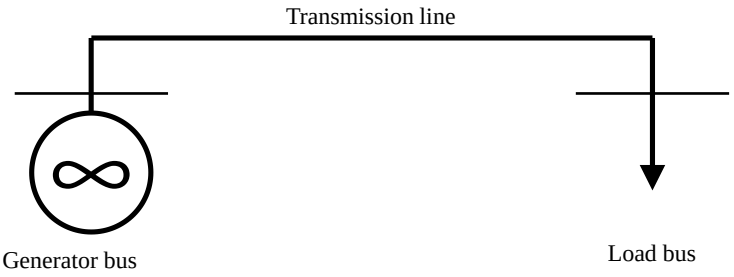


Figure 1: A single-load infinite bus system.

Figure 1 shows a single-load infinite bus system consisting of a generator of infinite capacity, a transmission line and an arbitrary load. The same system is represented as a per-phase single-line diagram showing equivalent impedances of the transmission line and the load.

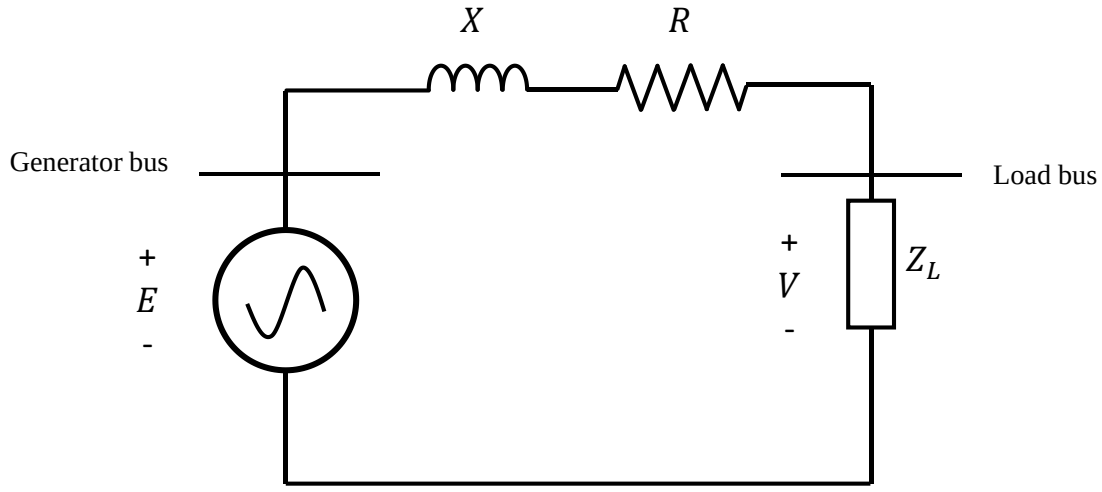


Figure 2: Per-phase representation with equivalent impedances

The maximum power problem is a classical electrical engineering problem and can be found in any introductory primer on circuit theory. The basic problem statement is that, given a system

with an internal ideal impedance, $Z_{Th} = R + jX$, and an internal ideal voltage source, E, there is an absolute maximum amount of power that can be transferred to an external load impedance, Z_L. It can be shown that the power consumed by the load Z_L is maximized when the load impedance is equal to the complex conjugate of the system impedance, i.e. when $Z_L = Z_{Th}^*$.

With some additional derivation, it can also be shown [1] that under a constant power factor, the power consumed by the load is maximized when the magnitudes of the load impedance and the system impedance are equal, when $|Z_L| = |Z_{Th}|$. Conceptually, this property is illustrated in Figure 3 below.

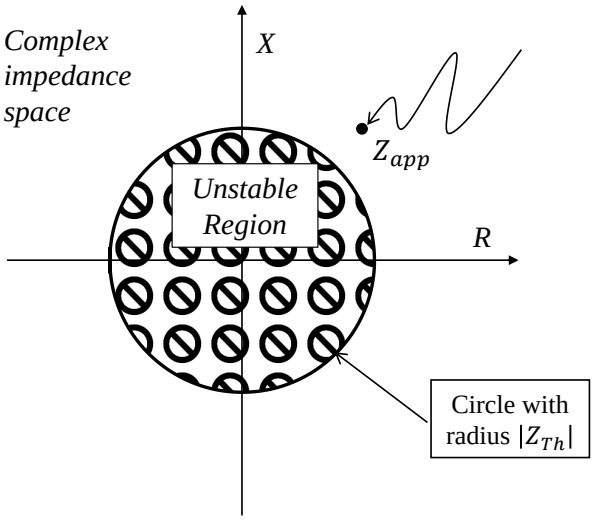


Figure 3: Load impedance approaching the off-limits region in complex impedance space

The family of P-V curve in Figure 4 illustrates the power output and the corresponding voltages from the load flow solutions. Note that for any given power output, there are exactly two, one, or zero voltage solutions when the power factor is held constant.

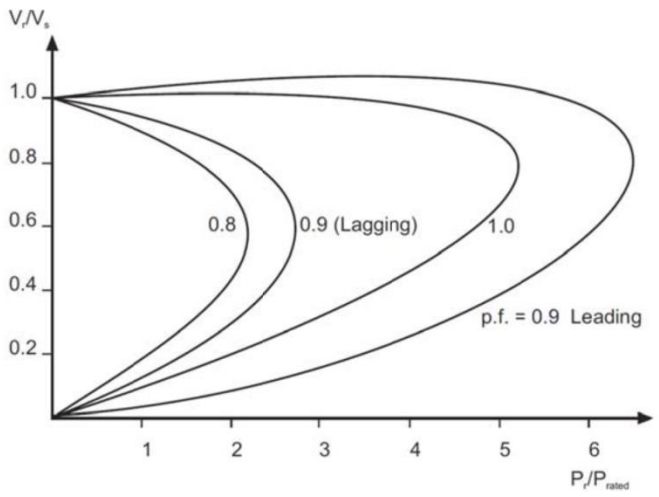


Figure 4: P-V Curves at various power factors Reprinted from [1]

The no solution case domain, to the right of the curve, represents power levels beyond the maximum power transfer point which are not physically possible. The point at the tip of the curve represents the maximum power transfer case discussed above. In power systems literature, this is often referred as the bifurcation point since, to the left of this point, there are two possible solutions for any power level delivered to the load. This point is also the stability limit of the system since, to the right of it, there are no solutions. The upper region of the curve is the stable region; the system is operating as designed and the voltage is relatively stable. Since current is related to real power by $P = VI\cos\theta$, at a given power factor (constant θ) the higher voltage corresponds to a lower current. The lower region represents the unstable region. At any point on in this region the power system will tend to deteriorate. The lower voltage solution requires more current, which will lead to equipment overloads, generators automatically tripping and constant power loads demanding more current. Since the demands cannot be met, without corrective action, the system will deteriorate until it collapses.

Figure 5, bellow, illustrates two ways in which a system can enter voltage collapse. In the first illustration (Figure 5a), the load increases until it has surpassed the loadability limit of the system. In the stable region, as the load demand increases, the equivalent load impedance, Z_L, decreases. Once the stability limit has been reached, further decrease in Z_L results in higher current but lower power delivered. In the second illustration (Figure 5b), a disturbance changes the characteristic of the system so that the load that was previously sustainable is no longer a possible solution.

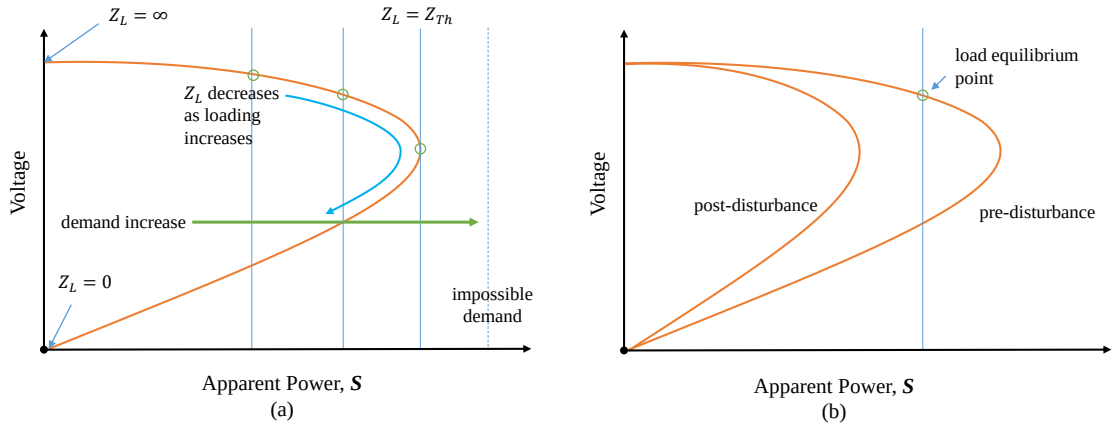


Figure 5: System conditions leading to voltage collapse. (a) The demand increases past the ability of the system to supply power. (b) System disturbance causes system stability limit to decrease below demand.

Voltage stability is defined by IEEE as the ability to maintain system voltage such that when the load admittance increases, the power delivered to the load will increase accordingly [12]. Such events are characterized by a progressive fall in voltages and a shortage of reactive power [13]. As the transmission system becomes ever more loaded, the risk of voltage collapse increases. In reality, the collapse will start with heavy loading (the situation illustrated in Figure 5a) followed by an unplanned switching action such as a generator trip, relays protecting an overloaded line, or a fault [14]. The culprits behind voltage instability are usually loads. Post-disturbance, loads

tend to be restored with a higher current demand than the pre-disturbance equilibrium, due to many motors spinning up, or loads controlled by thermostats, which normally operate intermittently all coming on at once. This increase stress on the network by increasing reactive power consumption causing further reduction in voltage. When the dynamics of the loads being restored attempt to consume power beyond the capability of the transmission system, the system can spiral into a state of voltage collapse [12].

As discussed above, the reactances of transmission lines restrict the flow of power since an increase in real and reactive power flow across lines also increases the voltage drop across the lines. In addition to the inherent limit due to the maximum power transfer, the capability to supply load is further reduced when generators hit their field or armature limits or current limits. Once this happens the generators effectively "run out" of reactive power and lose their ability to control the voltages at their terminals. Mathematically, this means that the voltage at the bus is no longer and independent variable in the load-flow equation set. When the generator is able to control power (P) and voltage (V) at the bus it is known as a PV bus. Once the generator can only inject power (P) and a constant reactive power (Q) the bus is known as a PQ bus. This is referred to as a PV-PQ transition. From a power-flow solution perspective, the generator has the same properties as a load bus except that power is injected rather than delivered, and both are referred to as PQ-type buses.

In practice, methods for voltage collapse protection involve a combination of planning studies, centralized system monitoring and simulation. For example, in the Texas grid, the Electric Reliability Council of Texas (ERCOT), which is the regional system operator and reliability coordinator, protects against voltage instability through a multi-stage approach [15]. At the planning stage, system changes such as new generation, unit retirements, or infrastructure

upgrades are studied regularly. In these studies, planning engineers model adverse system conditions and perform simulations to determine if system changes might lead to instability. If a potential for instability is found, a Generic Transmission Constraint is developed that is used in operations as an additional constraint for determining real-time dispatch. Since these studies are performed far in advance, they are not able to account for real-time system conditions, such as the availability of reactive power or outaged equipment that will have a direct effect on the loadability of the system. To account for this, a simulation is performed hourly that finds voltage collapse limits by increasing power transfer to specific areas of the grid until collapse occurs. Since the simulation is performed hourly, it is able to consider current system conditions to make a more accurate assessment of the system's proximity to a voltage collapse event. Additionally, in areas that are known to be vulnerable to voltage collapse, under-voltage protection relays are installed by the transmission operators which can react quickly to shed load.

Voltage collapse events tend to be relatively slow system events occurring over a period of minutes or hours. [13]. This long-term voltage stability typically involves slow-acting equipment such as tap-changing transformers and generator current limiters [12]. However, short-term voltage stability responsiveness is needed since fast acting components, such as HVDC converters and electronically controlled loads, can lead to rapidly occurring instability as well. Additionally, as the grid moves away from large synchronous machines towards fast-acting inverter-based resources, the need for fast response to voltage stability is likely to increase. Undervoltage load shedding relays that are typically employed as a mechanism to quickly respond to potential voltage instability pose a few problems. Choosing a setpoint is problematic because voltage is often a poor indicator of instability. Thus, a fixed set-point for relay operation may lead to unnecessary shedding or, worse, failure to operate when voltage instability occurs.

[16, 17]. The need for fast-acting voltage stability analysis leads to the use case for the VIP device or a similar fast-acting device, discussed in the next section.

1.2.1 Voltage Instability Predictor (VIP)

Originally described as the "Stability Monitoring and Reference Tuning Device (SMARTDevice)" [18], the Voltage Instability Predictor (VIP) uses an estimated Thévenin equivalent of the system to make predictions about the proximity to voltage collapse. This technique only requires local measurements of bus voltage and current to make its estimates.

VIP devices make use of the maximum power transfer theorem to track the proximity of a system to voltage collapse. These devices measure local voltage and current phasors and use these measurements to approximate the system as a two-bus network using Thevenin principles. Since the VIP device uses only local measurements, it does not rely on a system-wide communications infrastructure to make decisions about voltage collapse proximity (margin). Additionally, the simplicity of the calculation and possibility of automatic actions means that the device is capable of responding faster than other methods of protection against voltage collapse. While VIP might not be a replacement for centralized system assessment, in the event of an unforeseen voltage collapse event, it would be able to respond faster than operating instructions from a centralized system operator. Indeed, the inventors of the device propose two complementary roles for such a device: 1) it could be used as an alert system when it detects locally weak system conditions and 2) it could operate as a relay as it encounters more severe conditions[19].

The estimation of the Thevenin parameters is key to determining the stability margin at a measurement location. Generally, once two or more measurement sets are available, the

estimation can be performed [20]. Researchers in [21-23] use a recursive least-squares (LS) method to estimate the parameters, while other researchers employ Tellegen's theorem to derive an equation for estimating the parameters using two sets of operating points. The advantage of this approach, along with being computationally simpler, is that the estimated parameters are faster to update if a dynamic event occurs as opposed to LS methods, which require a significantly large data window to suppress oscillations. Other more creative and complex methods of estimating Thevenin parameters are examined in the literature review of this manuscript. The method used primarily through the course of this research is based on the approach using two operating points and is most succinctly described by the authors of [19]. The Methodology section (2) and appendices of this report gives an overview of the computations involved.

One of the main advantages of a VIP device is that it only uses local measurements and therefore doesn't require system-wide knowledge to operate. However, the idea of using networked VIP devices to improve stability margin estimates has also been proposed by various researchers [24, 25]. Combining evidence from multiple VIP devices implies some form of knowledge beyond its own observations. This is not a contradiction because a communication scheme between VIP devices doesn't mean that each device would need total system knowledge, requiring large bandwidth, computational power and communications infrastructure investment. If nearby VIP devices can be used to improve the predictive accuracy or the confidence of the prediction, then the net benefits of the non-centralized approach to voltage stability prediction are preserved and only a minimal communication network would be needed. Communications might be limited to a very narrow set of data such as the predicted stability limit along with a measure of the device's confidence in its predictions. The localized and autonomous nature of the VIP device would thus

be preserved. It is in this vein that the research presented here attempts to demonstrate distributed algorithms whereby neighboring VIP devices might work together for a more accurate solution.

Depending on the desired implementation scheme the VIP device could be used to add a layer of reliability to communication. System-wide communication could be through a less reliable medium, for example a fiber-optics link with a centralized control center. Communication with nearby VIP devices might be through a decentralized communication scheme which would be less vulnerable to a single point of failure. If system-wide communication is lost, the device can still operate.

1.2.2 VIP Literature Review

In [26] the authors solve the Newton-Raphson power flow problem at operating points, watching the Jacobian for zero-crossings. They observe that if there is no change in sign in a continuously variable iterative process and that the starting point of the iteration is stable, then the solution will also be stable. Since the iterative process is necessarily executed in discrete steps, the possibility of an unobserved double zero-crossing (while rare in practice) could falsely indicate a stable system. To counter these problems, researchers propose a modified Newton's algorithm which makes it possible to approach the point of voltage collapse with accuracy. In order to find a robust solution, the authors develop a technique that changes the iterative step size. In effect by decreasing the step of ΔP, according to the tolerance of the load flow solution, they determine the proximity of P to the stability limit and the last operating condition computed by their method will correspond to the stability limit. For large systems, this method could become quite time consuming, especially with the technology available at the time of publication in 1975.

In [27] researchers examine the problems of multiple load flow solutions when the solutions are similar (i.e. near the point of bifurcation) to determine which operating point is stable. In their study on a 6-bus system, they find the flat start solution to be more stable than the solution given by the exhaustive search. They also found that by examining the signs of the sensitivity matrix for voltage and reactive power control defined by [28], they are able to determine which solution is the stable one. In addition, by using the concept of stored energy in a system, they can make a similar determination based on whether the stored energy increases or decreases as frequency changes. They point out that the methods proposed by [26], which rely on the Jacobian solution method, have the disadvantage that it requires knowledge of a stable operating point, while the methods they propose using the sensitivity matrix and stored power concepts do not require this a priori information.

In [29] and [30], Lof et al. use singular value decomposition of the Jacobian to determine the stability limit. For a real symmetric matrix, the absolute values of the eigenvalues obtained in an eigenvalue decomposition are equal to the singular values of that matrix. The singular values are fairly insensitive to perturbations of matrix elements and are, thus, said to be well-conditioned. A fast algorithm for finding the minimum singular value is proposed.

Chiang et al. [31] delve into dynamic factors that influence the voltage stability problem. They assert that, in particular, classical load models such as constant P-Q, constant impedance, and constant current models are insufficient for capturing severe collapse dynamics. Researchers use the center manifold voltage collapse model to capture system dynamics after bifurcation and investigate how interaction between loads and generators may cause voltage collapse. The authors conclude that voltage collapse could be avoided by manipulating system parameters such that the bifurcation point is outside of the normal operating region.

Hawkins et al. [32] propose an algorithm for assessing the margin of a power system to the voltage collapse point. This margin is expressed as the additional active power transfer which could be sustained. The algorithm monitors the voltage sensitivity elements of the Jacobian elements, $\partial V_i/\partial P_i$ and $\partial V_i/\partial Q_i$. The researchers state that a sign change in the inverse Jacobian (which heralds the system collapse) can be the result of a generator reaching its reactive power limit or could occur when the voltage droops as the load increases. Their proposed method predicts the precise collapse point by the following: (1) After observing a sign change, the inverse Jacobian is recalculated with the most recently limited generator restored back to a voltage-controlled (PV) bus. (2) If the elements are still negative, then they linearly interpolate between the current point and the point of the previous PV-PQ transition.

Vu, et. al describe a system which uses only local measurements to predict the proximity to voltage collapse. This system calculates the Thévenin equivalent of the system as seen from a particular bus along with an equivalent impedance of the load. Modeling the system as this two-bus equivalent network, the system can be described by the equation,

$$V_L = E_{Th} - Z_{Th} I_L .$$

Since E_{Th} and Z_{Th} are not directly observable and are complex numbers, the system contains four unknowns and is not directly solvable. Vu, et al. propose the use of a least-square algorithm to find the best solution based on this equation with multiple measurements. The system described by these researchers will become known as the Voltage Instability Predictor (VIP) which forms the basis for much of the research contained in this study.

A difficulty of the VIP method pointed out by Corsi et al. is that the equation (above) has infinite results unless the assumption is made that throughout the interval between two measurements,

E_{Th} and Z_{Th} remain constant. However, when measurements are taken too close together, the corresponding matrix equation has a high risk of being singular if there is no variation in observable parameters. An accurate result is only produced when there is significant system variation, which often happens too close to collapse for any meaningful action to be taken. As such, Corsi et al. propose an algorithm to speed up the identification of Thevenin parameters. In this algorithm, the direction of change and the amount of variation in the Thevenin voltage are taken into account.

1.3 Sensor Fusion

Sensor fusion is a class of data analysis techniques that use multiple data sources in a similar way that humans' cognitive processes make deductions based on observations about the environment [6, 7]. We are constantly performing a form of data fusion in our daily lives, updating previously held beliefs based on new data which may be incomplete or unreliable Sensory information in the form of sight, smell and touch combine to tell us if food is delicious or rotten. Since information from one sensor might be incomplete or inaccurate, sensor fusion is an attempt to take the advantage of the multiplicity of data to fill in these gaps. Applications are widespread ranging from image processing [8], medical imaging, target tracking and identification [9], and robotics and artificial intelligence[10].

More formally, sensor fusion is the “acquisition, processing and synergistic combination of information” [7] gathered by a multiplicity of sources with the aim of providing a better understanding of the phenomenon being examined. If successfully implemented, sensor fusion can offer the following benefits when compared to data from only one sensor or multiple sensors examined separately:

- —multi-sensor systems are inherently redundant. With data available from multiple sources, the system can operate with the failure of one or more sources, albeit with reduced performance. From a power systems perspective, this is important because planned and unplanned outages often occur.
- Coverage area—multiple sensors can be arranged over a wide area to allow observations for a region larger than what any single sensor could observe. For making observations about system voltage stability, this is important because voltage collapse is a system-wide phenomenon.
- Confidence—sensors that agree can confirm each other's observations, increasing confidence in the fused observation.
- Response time—since more data is collected simultaneously, observations can be made with high confidence in a shorter time interval than they might be from one sensor collecting data alone.
- Accuracy—multiple sensors observing the same phenomenon can result in a more accurate observation than any single sensor could.

Techniques for sensor fusion include sequential estimation methods such as Kalman filtering, pattern recognition techniques based on clustering algorithms or neural networks, and decision-based techniques such as Bayesian inference or Dempster-Shafer's method [9]. Classical inference gives the probability that an observation can be assigned to a particular object or event given a hypothesis if the probability distribution function (PDF) is known. However, classical inference has some difficulty when trying to fuse different events [33]. It is difficult to obtain the probability density function of the observable parameter. Furthermore, classical inference has the ability to assess only two hypotheses at a time: the test hypothesis and the null hypothesis.

1.3.1 Dempster-Shafer Evidential Theory

Dempster-Shafer (D-S) evidential theory allows sensors to contribute detection or classification information to the extent of their knowledge. D-S theory can be interpreted as a generalization of probability theory in which probabilities are distributed to sets of propositions instead of mutually exclusive singleton propositions [11]. These probabilities are summed together as "Belief" functions, annotated $Bel(X)$, which represent the total confidence that can be placed in the proposition or subset. The set of all propositions under consideration are called the "focal frame", which sums to

Uncertainty problems can be classified into two types of uncertainty: aleatory uncertainty and epistemological uncertainty [11]. Aleatory uncertainty results from the stochastic nature of systems, in other words, randomness. This type of uncertainty can be dealt with through the frequentist approach to probability. Given a set of data with random noise and a sufficiently large sample size, we can construct a probability distribution function to describe the expected result of a parameter. The second type of uncertainty, epistemic uncertainty, arises from lack of knowledge about a system. This is a property of the observers (which will frequently be referred to as sensors in this study) and not of the system itself.

If an observer (sensor) reports based on its observations, that it believes with a 70% confidence that it has witnessed event A, does this imply that it has 30% confidence that it did not observe event A? Or does it simply mean that it has 30% confidence that it does not know what it was a witness to? Moreover, are these confidence assignments based on statistical data? Or are they based on a heuristic using imperfect information? If an observer believes it witnessed A with 100% certainty, does that mean that it is 100% certain that A happened, or can the observer be

wrong? Can we ever really have complete knowledge of anything? Epistemological musings aside, these are types of questions which traditional probability struggles to answer and D-S theory attempts to provide a framework for by incorporating two main concepts: The Theory of Belief and the Combination Rule.

These types of questions were pondered by Bernoulli[2] in the foundational text on probability, *Ars Conjectandi*. Bernoulli distinguishes two types of arguments: "pure" and "mixed" [34]. A pure argument can prove something in such a way that it does not prove anything in other cases. A mixed argument proves something in such a way that it will prove the contrary in other cases. He provides an example illustrating these argument types concerning the misfortune of Gracchus, which is here annotated with equivalent D-S Belief functions following Yager and Lui[35]. Suppose a person was stabbed in a crowd and reliable witnesses say that the perpetrator was wearing a black cloak. Gracchus and three others in the crowd were observed to be wearing a black coat. Upon subsequent questioning, Gracchus pales. The fact that Gracchus is wearing a black coat is a mixed argument. It provides evidence both for his guilt and for his innocence (if he wasn't wearing it). Additionally, we can write the belief function for this argument. Since there are four total suspects wearing a black coat and only one of the suspects is the murderer, the amount of belief it contributes to his guilt is $Bel(guilty) = 1/4$ and $Bel(innocent) = 3/4$. Since he must be either guilty or innocent, the belief in the set of all propositions, $\theta = \{guilty, innocent\}$, is $Bel(\theta) = 1$. This belief function is additive since all beliefs sum to 1. Gracchus' pallor is a "pure" argument. Since he might have turned pale for any number of

[2] Interestingly, not the Bernoulli who developed Bernoulli's principle of fluid dynamics, but his uncle.

reasons (guilt, sickness, cold, fear, anger, etc.). It only proves his guilt if he pales from a guilty conscience, but it does not prove his innocence if he pales for another reason. Thus, the evidence contributed by this observation $Bel(guilty) < 1$ and $Bel(innocent) = 0$, while $Bel(\theta) = 1$, as before. The beliefs no longer add to 1, thus $Bel(guilty)$ and $Bel(innocent)$ cannot represent probabilities that Gracchus is guilty or innocent.

A key feature of D-S theory is that it allows the assignment of a portion of a sensor's knowledge about a proposition to uncertainty instead of assigning it to negation of the proposition. This allows decisions about fused data which might be based on an incomplete probabilistic model. Because of its flexibility, D-S theory provides a powerful tool for multi-sensor fusion. For example, a sensor gathers evidence that suggest with a 10% confidence that it is detecting event A, with 50% confidence that it is detecting event B. Under probability theory, the remaining 40% confidence is assigned to the negation of A and B, in other words, the lack of evidence *for* either A or B is interpreted as evidence *against* A and B. Under D-S theory the remaining 40%, is assigned to uncertainty. These confidence measures are referred to in D-S theory as basic probability assignments (BPAs) or probability mass assignments. In this work, the term BPA is employed.

1.3.2 Basic Probability Assignments

D-S theory relies on BPAs to rate the amount of belief a sensor can contribute to the optimal hypothesis selection. Even if a sensor ascribes low BPA to an observation, the observation can still turn out to be the best hypothesis if the sensor ascribes high BPA to its uncertainty (or if it is outvoted by its peers). The BPA assignment of set A can be referred to as the probability mass of A and written $m(A)$. It defines a mapping of each set within the power set $\mathbb{P}$ (the set of all

subsets of the range of hypotheses) to a value between 0 and 1. The value that is assigned to $m(A)$ is the support for the evidence of only the set A and does not include the BPAs of any subsets of A. The criteria for BPA can be formally defined by the following relationships [11]:

$$m: \mathbb{P}(X) \rightarrow [0,1]$$

(BPA for every proposition is a value between 0 and 1.)

$$m(\emptyset) = 0$$

(The BPA assigned to the empty set is 0.)

$$\sum_{A \in \mathbb{P}(X)} m(A) = 1$$

(The BPAs of every subset of propositions within the set of all propositions must sum to 1. It's still probability, after all.)

A given set of propositions under examination is often referred to as the *focal frame* and notated by θ (not to be confused with the empty set symbol, $\emptyset$). Since the focal frame is a subset of itself, BPA can be assigned to the focal frame just like any other subset. The focal frame plays and important role in D-S theory because it represents the lack of evidence towards any specific proposition or proper subset of the focal frame. This ability to assign BPA to a lack of knowledge is central to D-S theory and plays a key role in the data fusion mechanism.

1.3.3 Theory of Belief Functions

D-S theory uses the concepts of Belief and Plausibility as a way of contextualizing uncertainty. The belief function arises from assigning probabilities to a set of propositions rather than to

individual points. Belief in a hypothesis does not measure the probability that it is correct, but rather the sum of the BPAs of the evidence that are in support of the proposition. Belief is defined as the total BPAs assigned to a subset of compatible propositions. If A is the set of compatible propositions and B represents subsets of those propositions, then

$$Bel(A) = \sum_{B|B\subset A} m(B)$$

The upper bound of the probability interval, Plausibility (Pl) is the sum of all BPAs of the sets that intersect the set of interest. Using the same notation with sets A and B, the expression for Pl can be formally written as

$$Pl(A) = \sum_{B|B\cup A\neq\emptyset} m(B)$$

Plausibility can also be defined by the complimentary relationship, which is the total confidence not assigned to the negation of Belief [36].

$$Pl(A) = 1 - Bel(\bar{A})$$

Historically, the Belief and Plausibility functions define the lower and upper bounds of the probability interval, as mentioned above. These functions were initially referred to by Dempster as the upper and lower probabilities in his foundational work on the topic [37]. Within this framework, the range defined by $\big(Bel(A), Pl(A)\big)$ bounds the "true" probability of *A*. In other words, if *A* is a subset of events with $P(A)$, then $P(A)$ should fall within the range bounded by the *Bel* and *Pl* functions.

Shafer's interpretation abandons the concept of lower and upper probability in favor of a semantic interpretation: The "belief" that we have in a hypothesis is the amount of "evidence" that we can find in support. The "plausibility" of a hypothesis is defined by the lack of evidence against it (but not necessarily for it). In other words, Shafer[36] viewed belief in a hypothesis not as a measure of the probability that it is true or not, but rather of the strength of the arguments that are in support of it.

The applicability of belief functions can be boiled down to three separate uses when applied to problems containing multiple or incomplete sets of information[38]:

1) *Representing incomplete knowledge*. Belief functions can be used to represent partial knowledge in cases where information is incomplete. When Bel(A) is interpreted to mean the strength of the evidence in favor of A and the Pl(A) is interpreted to mean the lack of evidence against A, then the probability masses or BPAs can be assigned as $m(A)$ (probability of arguments for A) and $m(\bar{A})$ (likelihood of arguments against A) where, crucially, $m(A) + m(\bar{A})$ is not necessarily equal to 1.
2) *Belief updating.* Belief functions provide a method of incorporating new evidence into partial or incomplete knowledge of a state. The method for updating a $Bel(A_1)$ with $Bel(A_2)$ for a new combined function, $Bel(A_{12})$ is given by the orthogonal sum rule known as Dempster's rule, the procedure for which is explained in the following section.
3) *Pooling evidence from multiple sources.* When there is evidence available observing the same event from multiple sources, the individual bits of evidence can be combined using Dempster's rule as well. Every piece of evidence towards a proposition can used to update the belief function and reduce uncertainty. Collecting evidence from multiple

sources also requires a way to handle conflicting information. Support for conflicting evidence is assigned to the measure of uncertainty. Dempster's rule then uses renormalization to redistribute the weights to compatible propositions.

In this study, the hypotheses are formulated by asking which measurement location is best for determining the stability margin. Since only one location can be the "best," the propositions represented by these hypotheses are incompatible (non-overlapping). Therefore, any belief in a proposition is equivalent to the BPA assigned to the corresponding sensor. The methodology for choosing the BPAs at each measurement location is discussed in Chapter II.

1.3.4 Combination of Evidence

Since D-S theory often deals with multiple evidence sources, there needs to be a formal rule for combination of evidence. This rule is called Dempster's rule which calculates the aggregate of evidence given by BPAs from two sources.

The general form for the combination of evidence from two sensors which assign BPAs m_1 and m_2 onto a set of propositions A and B, respectively, is the orthogonal sum of the BPAs which is often written with the notation, $m_{12} = m_1 \oplus m_2$. The combined set of evidence, $m_{12}(C)$ is the sum of the product of the BPAs of compatible (intersecting) propositions: $\{C|A \cap B = C\}$. The sum is multiplied by a normalizing factor, K, which is calculated from the inverse of the one minus the sum of incompatible propositions, or when the intersection of the proposition set A and proposition set B results in the empty set. The resulting equations are

$$m_{12}(C) = \left(\sum_{A \cap B = C} m_1(A) m_2(B) \right) K$$

where

$$K = \frac{1}{1 - \sum_{A \cup B = \emptyset} m_1(A) m_2(B)} \ .$$

For larger values of K, there are fewer compatible propositions, and the overlap between A and B contains less information. $K = \infty$ implies complete contradiction between A and B.

In order to understand the fusion algorithm and explore its implications, it is instructive to consider a simple example using concepts which are already familiar. In this scenario, medical doctors, Dr. Alice and Dr. Bob, are attempting to diagnose a patient who is experiencing certain symptoms. For this example, there are two possible mutually-exclusive conditions, Condition X and Condition Y. Since there are two propositions, the focal frame is $\theta = \{X, Y\}$

Dr. Alice is 90% confident that she has seen the condition before, but thinks it equally likely to be a result of condition X or condition Y. Since, she is confident that she knows that the cause is condition X or Y, she assigns only a 10% to proposition that she does not know what condition is the root cause ($m_A(\theta) = 0.1$). The rest she splits evenly between condition X and condition Y ($m_A(X) = 0.45$ and $m_B(Y) = 0.45$). The assignment of BPA to θ does not imply 10% support that the condition is neither X or Y, but that a 10% belief that it could be either one or neither. While this seems to contradict the BPA rule in the previous section since, the support assigned to the empty-set ($\emptyset$) should be 0. However, there can easily be a 3rd unknown condition or set of conditions (call it condition Z) which would have 0 support and not change the outcome of the belief and plausibility functions that we use to describe Dr. Alice's diagnostic evaluation. D-S belief theory allows us to make the following statements about Alice's views, pre-fusion.

Belief that condition A is true is 45%. This is apparent, since the only support for A comes from the BPA for A:

$$Bel_A(X) = \sum_{x|x \subset X} m(x) = 0.45$$

Where X is an element of the power set of θ:

$$x \in \mathbb{P}(\theta) \text{ and } \mathbb{P}(\theta) = \{\{X\}, \{Y\}, \{X, Y\}\}$$

Plausibility that condition A is true is 55%. Plausibility describes the lack of evidence or support for against. Since X and Y are mutually exclusive propositions ($X \cup Y = \emptyset$), the support for Y implies evidence against X, while support for θ is not evidence against any specific proposition ($X \cup \theta = X$):

$$Pl_A(X) = \sum_{x|x \cup X \neq \emptyset} m(x) = m(X) + m(\theta) = 0.45 + 0.1 = 0.55$$

Similarly, the belief that Dr. Alice the condition is one that she has seen before is 90%:

$$Bel_A(\{X, Y\}) = \sum_{x|x \subset \{X,Y\}} m_A(x) = m_A(X) + m_A(Y) = 0.45 + 0.45 = 0.9$$

The plausibility that it is either X or Y is 100%, since she has no evidence to the contrary:

$$Pl_A(\{X, Y\}) = \sum_{x|x \cup \{X,Y\} \neq \emptyset} m_A(x) = m_A(X) + m_A(Y) + m_A(\theta) = 0.45 + 0.45 + 0.1 = 1$$

Dr. Bob is rather uncertain. He suspects that the symptoms are caused by condition X, but is not familiar with condition Y. Since he has little faith in his response the assigns 80% to the

proposition that he does not know ($m_B(\theta) = 0.8$) and 20% to support of the diagnosis for condition X. Since he does not have any knowledge of condition Y, he assigns 0% ($m_B(Y) = 0$). This assignment to $m_B(Y)$ does not mean that he thinks that there is a 0% chance that condition Y is the correct answer, he simply has no evidence to support this (and only this) proposition.

Using the same equations applied to Dr. Alice, we can make the following statements about Dr. Bob's diagnostic beliefs:

$$Bel_B(X) = m_B(X) = 0.2$$
$$Pl_B(X) = m_B(X) + m_B(\theta) = 1$$

$$Bel_B(Y) = m_B(Y) = 0$$
$$Pl_B(Y) = m_B(Y) + m_B(\theta) = 0.8$$

$$Bel_B(\{X, Y\}) = m_B(X) + m_B(Y) = 0.2$$
$$Pl_B(\{X, Y\}) = m_B(X) + m_B(Y) + m_B(\theta) = 1$$

There are two takeaways from these results: 1) the plausibility of condition Y is non-zero even though it was assigned no BPA and 2) the addition of a proposition with no BPA (in this case, Y) does not affect the outcome of the other propositions. If condition Y was left out of the focal frame, then we would still get the result that $\{Bel_B(X), Pl_B(X)\} = \{0.2,1\}$.

In order to perform the fusion operation, it is helpful to make a reference table which shows which propositions are compatible (their union results in result in a non-empty set) and which are incompatible (their union results in empty set),

	X	Y	θ
X	$X \cup X = X$	$X \cup Y = \emptyset$	$X \cup \theta = X$
Y	$X \cup Y = \emptyset$	$Y \cup Y = Y$	$Y \cup \theta = Y$
θ	$\theta \cup X = X$	$\theta \cup Y = Y$	$\theta \cup \theta = \theta$

and to summarize the BPA assignments,.

	Dr. Alice	Dr. Bob
$m(X)$	0.45	0.2
$m(Y)$	0.45	0
$m(\theta)$	0.1	0.8

The fusion algorithm is best visualized (and easiest to perform) in matrix fashion where each element of matrix M_{ij} is the product of $m_A(i)m_B(j)$. The last row and column are reserved for the BPA assigned to the focal frame. This approach is also convenient since it can be performed as multiplication of a row and column matrices given by the sets of BPA assignments. The resulting table is below.

	$m_B(X) = 0.2$	$m_B(Y) = 0$	$m_B(\theta) = 0.8$
$m_A(X) = 0.45$	$m_A(X)m_B(X) = 0.09$	$m_A(X)m_B(Y) = 0$	$m_A(X)m_B(\theta) = 0.36$
$m_A(Y) = 0.45$	$m_A(Y)m_B(X) = 0.09$	$m_A(Y)m_B(Y) = 0$	$m_A(Y)m_B(\theta) = 0.36$
$m_A(\theta) = 0.1$	$m_A(\theta)m_B(X) = 0.02$	$m_A(\theta)m_B(Y) = 0$	$m_A(\theta)m_B(\theta) = 0.08$

Using $\bar{m}$ to denote the un-normalized BPA after fusion, we can calculate,

$$\bar{m}_{AB}(X) = \sum_{x \cap y = X} m_A(x)m_B(y) = m_A(X)m_B(X) + m_A(\theta)m_B(X) + m_A(X)m_B(\theta)$$
$$= 0.09 + 0.02 + 0.36 = 0.47$$

$$\bar{m}_{AB}(Y) = \sum_{x \cap y = \mathrm{Y}} m_A(x)m_B(y) = m_A(Y)m_B(Y) + m_A(\theta)m_B(Y) + m_A(Y)m_B(\theta)$$

$$= 0 + 0 + 0.36 = 0.36$$

$$\bar{m}_{AB}(\theta) = \sum_{x \cap y = \theta} m_A(x) m_B(y) = m_A(\theta) m_B(\theta) = 0.08$$

As expected, the total BPA of the un-normalized probabilities does not sum to 1. To solve this, we employ Dempster's normalization factor, *K*, which redistributes the BPA that has been assigned to the empty set after fusion.

$$K = \frac{1}{1 - m_A(Y) m_B(X) - m_A(X) m_B(Y)} = \frac{1}{1 - 0.09} = 1.0989$$

After normalizing, the fusion results are

$$m_{AB}(X) = 0.516$$

$$m_{AB}(Y) = 0.396$$

$$m_{AB}(\theta) = 0.088$$

$$Bel_{AB}(X) = 0.516$$

$$Pl_{AB}(X) = m_B(X) + m_B(\theta) = 0.604$$

$$Bel_{AB}(Y) = m_B(Y) = 0.396$$

$$Pl_B(Y) = m_B(Y) + m_B(\theta) = 0.483$$

$$Bel_B(\{X, Y\}) = m_B(X) + m_B(Y) = 0.912$$

$$Pl_B(\{X, Y\}) = m_B(X) + m_B(Y) + m_B(\theta) = 1$$

1.3.5 Handling Conflicting Data

As a number of researchers [11, 38-40] and theorists have pointed out, handling datasets with large degrees of conflict can be problematic and produce counter-intuitive results. Of course, the purpose of performing fusion is to incorporate various sources of evidence which disagree to a

certain extent. If there is no conflicting evidence, there would be no need for fusion as every source of information would be complete and correct. In case of complete disagreement in which there is no uncertainty, the Dempster's rule calculation is undefined. Suppose that $m_1(A) = 1$ and $m_2(B) = 1$. The fusion would result in $m_{12}(A) = m_{12}(B) = \frac{1}{0} \cdot 0$. Since all BPA would be assigned to empty-set propositions (100% disagreement) and no BPA would be assigned to non-empty sets (no agreement), the normalization factor is $K = \frac{1}{1-1}$. The failure to find a solution is a shortcoming, but this result is not necessarily count-intuitive. After all, without being provided with any other information, any algorithm aimed at reconciling two completely contradictory pieces of information will fail to provide meaningful results.

Yager [40] demonstrates how the D-S fusion algorithm can produce results which are indeed counterintuitive under high degrees of disagreement. Given a scenario with three propositions over the frame $\theta = \{a, b, c\}$ and two sensors reporting the belief assignment sets m_1 and m_2, the following results are reported:

$$m_1(a) = 0.99 \qquad m_2(b) = 0.01$$

$$m_1(b) = 0.01 \qquad m_2(c) = 0.99$$

The resulting output from the fusion algorithm will be:

$$K = 10^4$$

$$m_{12}(a) = 0$$

$$m_{12}(b) = 1$$

$$m_{12}(c) = 0$$

A rational observer examining the sensor outputs would likely conclude the opposite result: proposition *b* is the *least* likely to be true since both sensors agree with a high degree of confidence that *b* is not true. The complete contradiction between propositions *a* and *b* and the resulting high normalization factor lead to all BPA being assigned to *b*. However, in this case the rational observer would not be able to determine which proposition is correct since the evidence for *a* and *c* is equally strong.

In a similar vein, a single faulty sensor, could cause results which are invalid even in a system containing many sensors which are in strong agreement. Given a three-sensor system with the focal frame $\theta = \{j, k\}$ the following outputs are reported:

$$m_1(j) = 0.99 \qquad m_2(j) = 0.99 \qquad m_3(k) = 1$$

$$m_1(\theta) = 0.01 \qquad m_2(\theta) = 0.01$$

The result of the fusion between m_1 and m_2 produces strong agreement as one would expect.

$$K_{12} = 1$$

$$m_{12}(j) = 0.9999$$
$$m_{12}(k) = 0$$
$$m_{12}(\theta) = 0.0001$$

However, further fusion with sensor 3, the faulty sensor, which is reporting 100% support for *k*, reverses the result:

$$K_{123} = 10^5$$

$$m_{123}(j) = 0$$

$$m_{123}(k) = 1$$

$$m_{123}(\theta) = 0$$

In this example, a ration observer or even some type of polling algorithm, would likely have easily concluded that *j* is most likely correct since two sensors have agreed on this proposition with a high degree of certainty. Furthermore, this example could be extended to an arbitrarily large set of sensors reporting a high level of agreement. Still, it would only require one faulty sensor reporting 100% confidence in the incorrect reading, to overrule every other sensor.

By now, the astute reader might have noticed the counterintuitive results obtained in the examples presented above manifest themselves when one or more sensors reports no uncertainty (i.e. assigns no BPA towards the focal frame, θ). If D-S fusion is an engine, then uncertainty is the oil that allows the parts to move and produce meaningful action. In Yager's example, adding even a small amount of uncertainty to the evidence gathered by the sensor provides entirely different—and much more reasonable—results. Instead of the sensor outputs defined above, consider a case where the mass assignments are modified to take 1% BPA from the proposition with the highest confidence and reassign it to the focal frame:

$$m_1(a) = 0.98 \qquad m_2(b) = 0.01$$

$$m_1(b) = 0.01 \qquad m_2(c) = 0.98$$

$$m_1(\theta) = 0.01 \qquad m_2(\theta) = 0.01$$

The new result is

$$K = 50$$

$$m_{12}(a) = 0.49$$

$$m_{12}(b) = 0.015$$

$$m_{12}(c) = 0.49$$

$$m_{12}(\theta) = 0.005$$

This result set intuitively makes sense with the input data. Support for *a* and *c* are equal and much larger than support for *b*, which is very small. By adding a small amount of structural uncertainty to the problem, we were able to produce reasonable results.

Another approach, which was developed for this study, is to divide every BPA assigned to a proposition by the total number elements in the focal frame before performing the fusion operation. This has the added benefit, that it allows the BPAs to be assigned arbitrarily to any value on the range [0 1] without having to worry about the total probability of a proposition set exceeding 1 (which would violate the rules establishing the concept of BPA). We can demonstrate this approach by modifying the "faulty sensor" example. Since there are two elements in the focal frame, the highest initial BPA that any proposition can be assigned is $m = \frac{1}{2}$. We obtain the following assignments:

$$m_1(j) = \frac{1}{2} = 0.5 \qquad m_2(j) = 0.5 \qquad m_3(k) = 0.5$$

$$m_1(\theta) = 1 - 0.5 = 0.5 \qquad m_2(\theta) = 0.5 \qquad m_2(\theta) = 0.5$$

Which results in an output of

$$K_{123} = 1.6$$

$$m_{123}(j) = 0.6$$

$$m_{123}(k) = 0.2$$

$$m_{123}(\theta) = 0.2$$

Now the support for *j* is the strongest as expected, but there is also a good bit of uncertainty as well as support assigned to *k*. This is reasonable given that there are only 3 sensors and one is in strong disagreement. Lacking any knowledge on the functionality of sensor 3, we have no evidence to suspect it other than its outlier measurement. It is still plausible (Pl = 0.4; Bel = 0.2), given the data that we have, that the two sensors in agreement are faulty. If we add additional sensors to the problem and they are in agreement with the first two sensors, the results will converge towards complete confidence in the proposition *j*. If we add two additional BPA sets, $\{m_4(j), m_4(\theta)\} = \{0.5, 05\}$ and $\{m_5(j), m_5(\theta)\} = \{0.5, 05\}$ the next two fusion results will be

$$K_{1234} = 1.111 \qquad K_{12345} = 1.059$$

$$m_{1234}(j) = 0.778 \qquad m_{12345}(j) = 0.882$$

$$m_{1234}(k) = 0.111 \qquad m_{12345}(k) = 0.0588$$

$$m_{1234}(\theta) = 0.111 \qquad m_{12354}(\theta) = 0.0588$$

The as additional sensors are added, the support for proposition *j* is strengthened beyond the support given by any one sensor and the effect of the faulty sensor are diminished. This approach forms the basis for the methodology employed in this study which is presented in Chapter II of this document.

1.3.6 Sensor Fusion Literature Review

Researchers, Xiao et al. [41], propose an iterative scheme, using the nodal average consensus to compute a maximum-likelihood (ML) estimate from distributed sensor data. Information is diffused across the network by updating each node with a weighted average from connected nodes. Local weighted least-squares estimates are then computed at each node. This approach avoids the difficulty of a centralized sensor fusion scheme wherein each sensor must send data to a data fusion center. In this distributed scheme, there is no single fusion sensor, and each node does not contain knowledge of the network topology. The researchers use the concepts of maximum-degree weights and Metropolis weights adapted from Markov chain construction methods when computing the distributed consensus. They show that after sufficient iterations, the individual nodes will reach consensus on a parameter (for example, a global average) that a centralized data fusion center would reach given complete knowledge of the system.

In [42], Olfati-Saber examines the problem of distributed Kalman filtering (DKF). Kalman filtering refers to a set of equations that provide an efficient recursive solution to the least-squares method, which was first described by R.E. Kalman 1960 [43]. Olfati-Saber proposes a method to divide the DKF problem into two separated dynamic consensus problems. The first involved decomposing the problem into a network of collaborative “micro-Kalman filters” with local communication. The second problem involves implementation of consensus filtering. He proposes a solution to the DKF problem which does not require a network with complete all-to-all links (which is rather impractical), as had been proposed by prior researchers. He employs a low-pass consensus filter for the fusion of time-varying measurements at a node and a band-pass consensus filter for the fusion of the inverse covariance matrices. The outputs of these filters feed into the distributed “micro-Kalman filters” which are able to collaboratively provide a state of

the observation. He concludes that the proposed DKF solution provides almost identical estimates to the centralized algorithm and is much more scalable.

He, et al. [44] use a modified form of DS theory to estimate the remaining useful life (RUL) of lithium ion batteries. Model parameters are computed by combining sets of measurements and sets of training data. The researchers record the capacity curves of three batteries and use a parametric fit to calculate model parameters and the 95% confidence intervals of each parameter. Dempster's rule is applied by first assigning the same probability mass to each parameter's "correctness," implying that each parameter estimate is equally credible. If a parameter estimate's confidence interval is observed to fall inside the confidence interval of another parameter's estimate, then the masses are summed. This is similar to Dempster's rule in which probability masses of subsets of a given set are summed during sensor fusion. Rather than selecting parameters based on the greatest probability, as is done in traditional DS fusion, the model parameters are calculated by weighted arithmetic mean with the resultant probability masses as the weights. Since the results rely on data-driven methods, the technique does not require specific knowledge of battery materials or properties to predict its RUL. He, et al. [44] then employ Bayes' rule with Monte Carlo sampling to update the parameter estimate after each fused observation.

A key problem in DS theory is the calculation of the mass function based on information provided by the sensors. As researchers in [45] point out, the definition of a probability mass assignment function remains a largely unsolved and the assignments are usually performed heuristically.

Researchers, Basir and Yuan [46], use Dempster-Shafer theory to help with engine fault diagnosis. The effectiveness of such a mechanism depends on how complimentary the information obtained by the sensors is. It is difficult to identify the correct results from data produced by sensors when it contains a wide array of contradictory information. Basir and Yuan [46] propose two methods for probability mass assignments. One such method is in terms of a feature vector which defines a vector of length *M* corresponding to measurements from *M* sensors. The measurements can be of mixed type. For example, one element may represent the RMS value of one sensor and another element could represent the kurtosis of the same sensor or a different sensor. If there are N fault states, an *N* by *M* matrix can be defined with each row corresponding to the known sensor outputs representative of each state. The inverse of the Minkowski distance between the measurement vector and the rows of the fault matrix are normalized and treated as probability mass assignments. The researchers are able to demonstrate that the fused information provides a more accurate engine fault diagnosis than individual sensors.

Xu, et al. [45] propose a method to determine probability mass assignments using properties of normal distributions from training data. They first perform a normality test and then a Box-Cox transformation as needed to ensure that the training data is normally distributed. From the transformed mean and standard deviation, a normal curve is obtained for each class in the data set. Thus, for each attribute, the set of normal distribution curves can be plotted for each class. For example, the researchers use a dataset which contains measurements for different types of iris flowers. One plot shows the normal distributions of sepal lengths for three iris species. Xu, et al. [45] show that the probability of a specific sample belonging to a class is proportional to the intersection point of the normal curve with the corresponding value of the sample. An algorithm

is developed that takes advantage of this property to assign probability masses. The researchers conclude that this is an effective method for probability mass assignment when training data is available.

2 METHODOLOGY AND SIMULATIONS

As discussed in the previous chapter, the simplicity of the VIP device provides advantages, but these advantages come at a cost: as shown in this study, the predictions made at each VIP sensor throughout the system tend to disagree. The purpose of this study is to take advantage of the multiplicity of data in order attempt to reconcile the data and make a better prediction about the actual stability limits of the system. With this goal in mind, in this chapter the modeling and implementation of the VIP devices are described. The techniques to combine evidence from multiple VIP devices are described as well.

2.1 Software

The simulations performed in this study are primarily performed in MATLAB using the MATPOWER v6.0 package. The MATPOWER package provides a suite of tools for solving power-flow cases in MATLAB. VIP modeling and data fusion methods were implemented using custom MATLAB scripts developed for this study. Additionally, Microsoft Excel was used for algorithm prototyping and ad-hoc data analysis.

2.2 Modeling the VIP Device

2.2.1 Thévenin estimation

In its most basic form, Thévanin's Theorem states that any two-terminal (i.e. single port) network of resistors and voltage sources can be simplified to a single resistor and voltage source. Under steady-state conditions, this theorem can be extended to include inductive and capacitive elements [47]. In this commonly used circuit simplification technique, the "Thévenin Equivalent" consists of an ideal Thévenin voltage source, E_{Th}, in series with an internal

Thévenin impedance Z_{Th}, and an applied external impedance Z_{app}. Figure 6 shows a system modeled as a Thévenin equivalent from the perspective of a load bus.

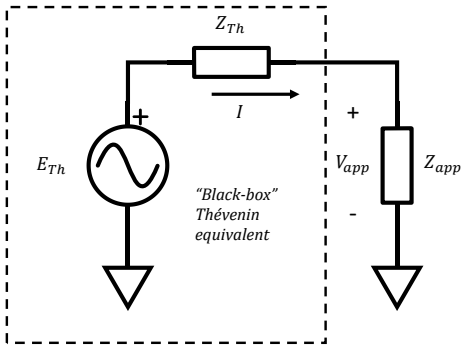


Figure 6: Thevenin equivalent model

Voltage collapse occurs when the power system infrastructure can no longer supply the power demanded by the load. The Thévenin equivalent is instructive because, it shows that voltage collapse is a simple phenomenon at its most basic level. It demonstrates that there is a physical limit to the amount of power that can be supplied to a load even if neither transmission nor generation limits have been exceeded. The components in an ideal Thevenin equivalent have no theoretical power limits of their own, yet there exists an inherent limit to the amount of power that can be transferred from the source to the load. As any student in an introductory circuits class can (hopefully) attest, this limit, the maximum power transfer, occurs in a purely resistive circuit when $R_{app} = R_{Th}$. While the derivation is a bit more complicated, the equivalent holds true for a reactive circuit: maximum apparent power ($\mathbf{S_{max}}$) occurs when $|Z_{app}| = |Z_{Th}|$.

The simplest way to determine the Thévenin parameters of a system is to measure the open-circuit voltage across the terminals and the short-circuit current with the terminals connected directly together. The Thévenin voltage is equivalent to the open-circuit voltage measured and

the Thévenin impedance is the ratio of the open-circuit voltage to the short-circuit current, i.e. $E_{Th} = V_{oc}$ and $Z_{Th} = V_{oc}/I_{sc}$.

In power systems, however, we do not have the luxury of being able to remove loads and/or short-circuit buses in the name of measuring the Thevenin parameters, so other methods must be used. Researchers discussed in the literature review have used a variety of methods. The most common are estimation by least-squares (LS) and a simple two-sample difference method.

2.2.1.1 Least-Squares Method

In this method of tracking Z_{Th}, researchers use new measurements to update the previous Thevenin equivalent values. This method was proposed in the original [18] "SMARTdevice" proposal for what is now referred to as a VIP device. The advantage of using this method is that it provides stable measurements that do not change wildly with new, potentially inaccurate, readings. However, decisions about whether or not a load is rapidly causing the system to approach voltage instability need to be made quickly, so the smoothing effect of the least-squares method isn't necessarily desirable.

While this method was explored in the preliminary research leading up to this study, it was ultimately discarded in favor of the Difference method (discussed in the next section). The L-S estimate does not respond quickly to sudden system changes and it requires significant processing power, which conflicts with the purpose of a distributed, fast-acting device.

2.2.1.2 Difference Method

The difference method is a simple method based on the premise that if the system is nearly constant, then the Thevenin parameters can be calculated from two measurements as the load is

varied. In practice, the measurements should be close enough together, so the system parameters do not have time to change significantly but are far enough apart that the load has time to vary meaningfully.

$$E = Z_{Th} I_1 + V_1$$

$$E = Z_{Th} I_2 + V_2$$

Equating these two equations and solving for Z_{Th} results in

$$Z_{Th} = \frac{V_2 - V_1}{I_1 - I_2}$$

Figure 7 shows trajectories of Z_L and Z_{Th} at bus 4 of the IEEE 39-bus test case as the system loading increases.

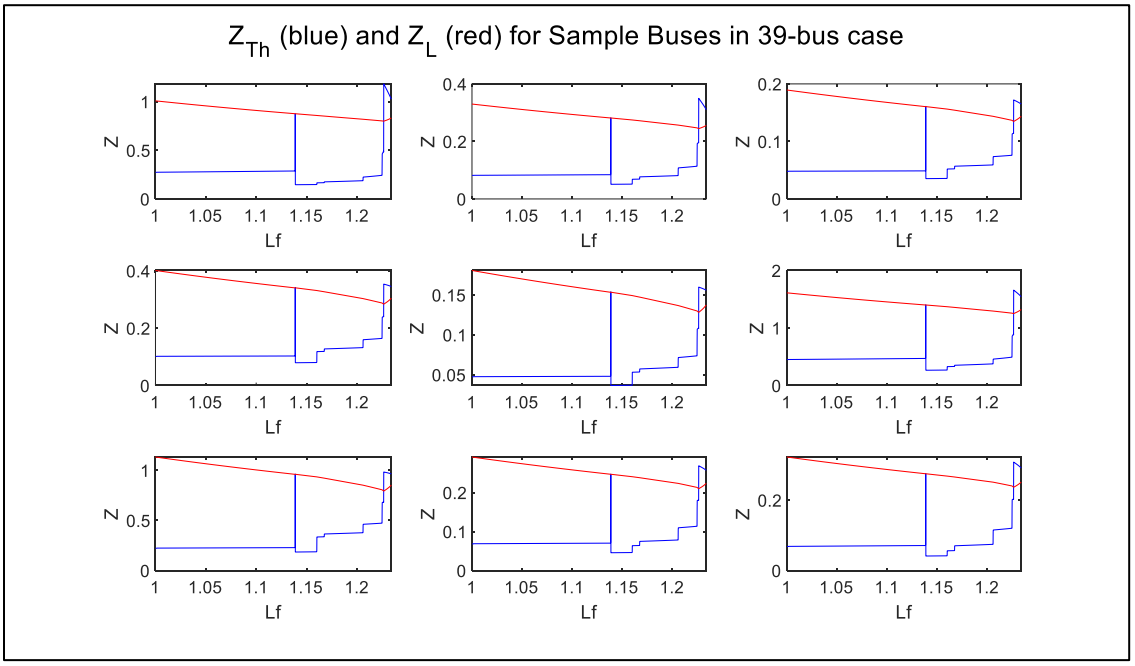


Figure 7: Thévenin impedance and load impedance as load factor increases for the first 9 load buses in the 39-bus case

2.2.2 Margin Estimation

2.2.2.1 Impedance Margin

Since the Thevenin-to-load impedance ratio is the marker for the point of maximum power transfer, the most natural way to represent the system stability margin is to directly describe the extent to which the impedances can change before collapse in terms of impedance itself. Simply put, the difference between the impedances is the amount of impedance "left" at a load bus before voltage collapse. This stability measure is referred to as the Impedance Margin. While this metric might be straightforward mathematically, it is not necessarily intuitive from a system operator's perspective as impedance is not a directly measured quantity like MW and MVA flows.

The margins can also be calculated from the difference between the measured load impedance and the Thévenin impedance. In this example, the maximum power is calculated from a linear extrapolation of two subsequent measurements of $\Delta Z_k = Z_{L,k} - Z_{Th,k}$.

Expressing the system loading factor at step k as λ_k, let λ_c be the point of collapse where $\Delta Z_{k=c} = 0$. In other words, when $k \to c$ then $Z_{L,c} \to Z_{Th,c}$. Using the point-slope form of a linear function, gives

$$\Delta Z_k - \Delta Z_c = \left(\frac{\Delta Z_k - \Delta Z_{k-1}}{\lambda_k - \lambda_{k-1}}\right)\lambda_k - \lambda_c \, .$$

When solved for the margin in terms of the load factor, $\lambda_c - \lambda_k$, and substituting $\Delta Z_c = 0$, the impedance margin is,

$$M_Z = \lambda_c - \lambda_k = \frac{\Delta Z_k}{\Delta Z_k - \Delta Z_{k-1}}(\lambda_k - \lambda_{k-1})$$

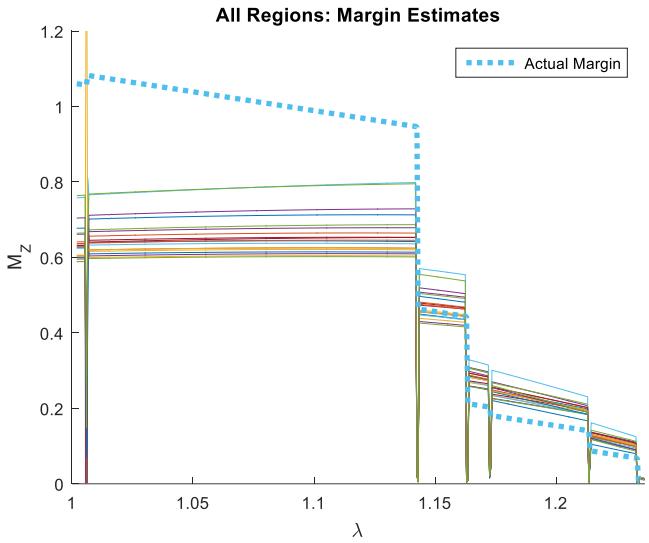


Figure 8: Impedance margin estimates on the 39-bus study case

The challenge of this method lies in estimating the impedance trajectories. In the above formulation a linear extrapolation is used. The trajectories can also be estimated by using a regression technique, and the coefficients of the resulting polynomial can be fitted by linear least squares or nonlinear least square techniques. The accuracy of the regression is impacted by the number of data samples used for estimating the curve. The Thevenin equivalent impedance is mostly constant, except when events such as PV-PQ transitions, topology-changing contingencies or generator trips take place.

2.2.2.2 Power Margin Method

In light of the difficulty conceptualizing the Impedance Margin, an alternative method is proposed by [19] and employed in this study which characterizes the system stability margin in

terms of apparent power, S. This Power margin, M_P, is defined as the difference between the power consumed at bus k and the maximum predicted power at that bus before voltage collapse.

$$\Delta S = S_D - S_{D,\max}$$

After Julian et al. the power margin is calculated using a linear extrapolation of the slope of the V-I curve calculated at a given iteration, k, to its maximum forecast [19, 23].

The slope of the V-I curve is calculated using the current and previous iteration's measurements. Incidentally, this is equal to the negative of Thevenin impedance:

$$\frac{V_k - V_{k-1}}{I_k - I_{k-1}} = -Z_{th}$$

A linear forecast satisfies the following relationship:

$$(V - V_k) = -Z_{th}(I - I_k)$$

Thus, the forecast power is quadratic with respect to I

$$S_D = VI = -Z_{th}I^2 + (Z_{th}I_k + V_k)I$$

The maximum will occur at the bifurcation point that occurs when the discriminant goes to zero

$$S_{D,\max} = \frac{(V_k + Z_{th}I_k)^2}{4Z_{th}}$$

Thus the difference is between the current an maximum power is,

$$\Delta S = \frac{(V_k + Z_{th}I_k)^2}{4Z_{th}} - V_k I_k$$

Which reduces to,

$$\Delta S = \frac{\left(V_j - Z_{th,j} I_j\right)^2}{4 Z_{th,j}}$$

Which is the power margin defined in [19].

For comparison across all buses, the margin is calculated in terms of the loading factor λ using the relationship, $\lambda_k \triangleq S_{j,k}/S_{j,0}$. Where $S_{j,0}$ is the power at bus j when $\lambda_0 = 1$. Thus the normalized power margin at bus k is

$$M_{P,k} = \frac{\Delta S}{S_{j,0}}$$

Figure 9 shows the results of the linear power margin estimation method applied to the IEEE 39 bus system. For comparison the true margin determined by simulating increasing loads until non-convergence is shown by the dotted line.

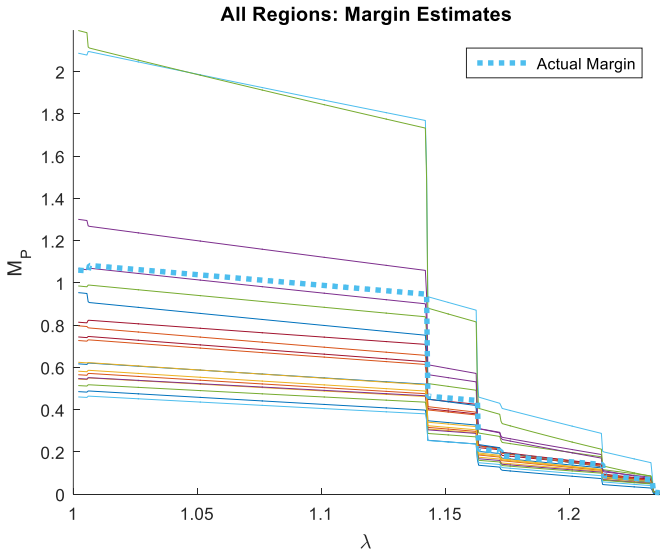


Figure 9:Power margin estimates on the 39-bus study case

2.3 Determining the Critical Load Factor

2.3.1 Continuation Power Flow

The MATPOWER package includes a continuation power flow (CPF) algorithm that, starting from a base scenario, uses continuation or branch tracing methods to determine steady state stability limits. The algorithm employs a predictor-corrector method to track voltage profiles at each bus as power increases. The program can be used to find the bifurcation point and establish a baseline maximum load factor (fig. Figure 10).

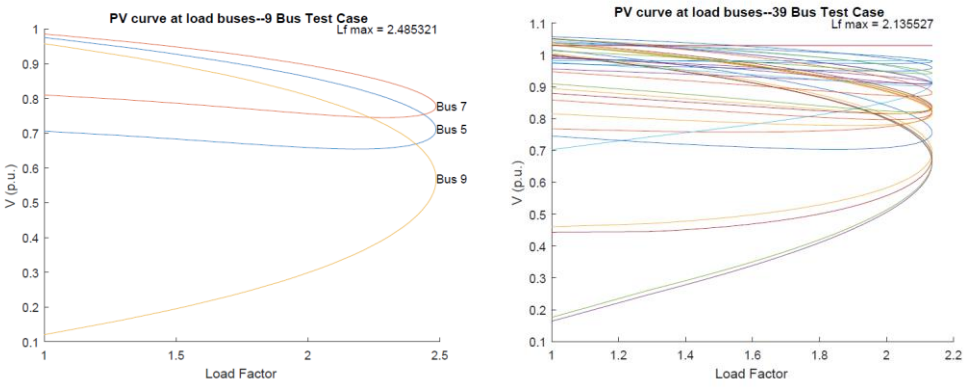


Figure 10: PV curves in the 9-bus and 39-bus test case calculated using CPF with reactive limits disabled

In the preliminary stages of this study, the CPF algorithm implementation in MATPOWER 5.0 had a shortcoming that hampered its usefulness to this study: it did not take into account PV-PQ transitions that occur when generator buses hit their reactive power limits. Even when a solution to the base case would include a PV-PQ transition for a solution, the CPF algorithm's implementation of the power flow solver ignores generator reactive limits.

Since MATPOWER 6.0, the CPF algorithm can process PV-PQ bus transitions that occur when generators reach their reactive limits. Therefore, this method can be used to accurately determine the maximum load factor.

2.3.2 Stepping Method

Due to the limitation of CPF at the outset of the study, alternate approaches were developed to find simulated system power margins when generator reactive limits are modeled. The simplest approach is to increase the load-factor and corresponding generator MW outputs by a small fixed per-unit amount and running N-R power flow until the system collapses. As the step sizes used in this study are usually rather small (0.0001 to 0.001) this process can be quite time-consuming, especially for large systems with buses that number in the 1000s. The maximum load-factor algorithm needs to be able to quickly determine the system limits since it will have to be repeated many times with different generator limits to determine the true margin. The algorithm for finding the true margin is described the next section (D). Using the results of the previous case rather than starting from the initial case when increase the load-factor improves the speed of this algorithm and simulates a realistic scenario since aggregate system load tends to increase incrementally in real-time. The drawback of this approach is that it is relatively time consuming, requiring many iterations to find the critical value.

2.3.3 Half-Interval Search Method

The algorithm employed in this study uses a binary search function or half-interval search function. This function works by incrementing the generation and load in the base case until the solution fails to converge. Once the first failure is encountered, the step size is halved and the generation and load are stepped back. If a convergence is found, the step size is halved, and the

power is incremented upwards. If the solution once again fails to converge, the step size is halved and the power is stepped backwards. This continues until the step size is smaller than the desired tolerance and the solution converges.

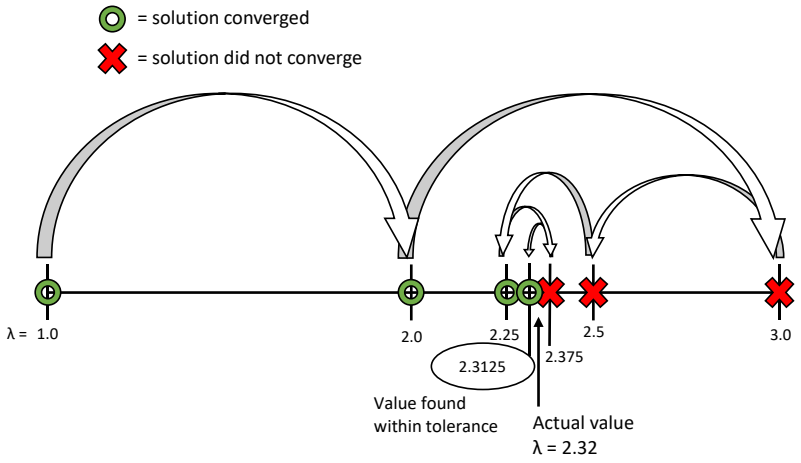


Figure 11: Half-interval search illustration

The advantage of this method is that it requires fewer N-R solutions, however it can result in a different solution than the stepping-method in the previous section since the steps between the solved cases are necessarily larger. For this reason, and since the CPF-based algorithm became available, this method was ultimately discarded.

2.3.4 Jacobian Zero-Crossing

An unanswered question in discussing the above methods for iterating towards the critical loading factor is what constitutes the system "failing to converge". Mathematically, this occurs when the Jacobian matrix becomes singular and can no longer be inverted. In a stepped iterative computation, it is not trivial to observe this condition exactly. Due to the iterative nature of the Newton-Raphson method, simplifying assumptions which are used to speed up the solution and inaccuracies that arise when inverting large matrices, the actual point of collapse where the

Jacobian is singular might not be observed directly. This can lead to the N-R solver finding un-physical solutions after point of collapse where the “knee” of the P-V curve bends towards unstable operation. A good example of this phenomenon is the IEEE 39-bus case that was examined in this study. As the loads in the system are increased the P-V curve (Figure 12, Figure 13) displays the characteristic voltage droop associated with the power transfer limit being reached but instead of the power flow solver failing, the voltages abruptly spike and continue to rise until the solution finally fails to converge. Clearly, this voltage spike is not a real solution, but rather a mathematical artifact of the solution method.

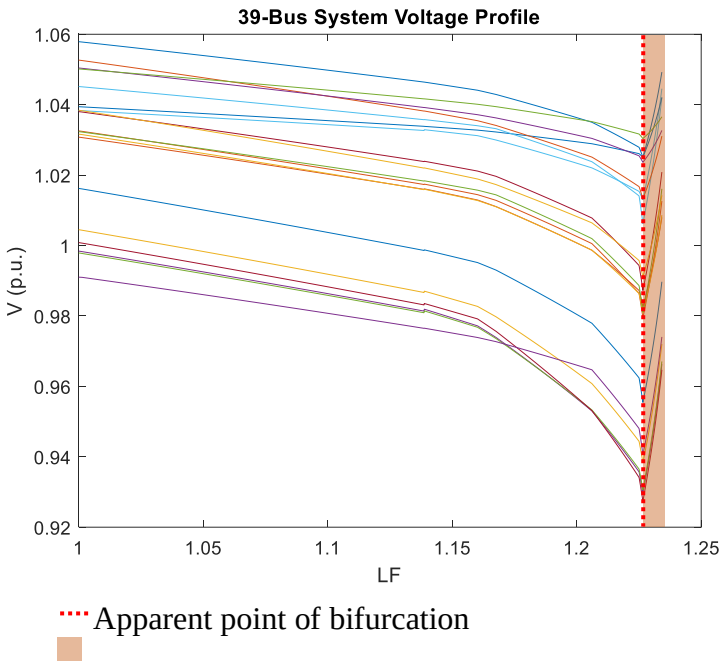


Figure 12: 39 bus case voltage profile

While a simple assumption may be that after X iterations, if the system does not converge, then it does not have a solution. However, this assumption is problematic, since a) the value if X is

arbitrary and b) the system may converge to a bad solution after sufficient iterations. Indeed, in this study, it was found that for certain cases, arbitrarily selecting a value of 10 max iterations vs 20 max iterations changed the maximum load factor determination. When performing power-flow analysis, especially of large, difficult to solve cases there are times when 50 iterations of the N-R algorithm might not converge, but 51 iterations might. For this reason, it was determined that using an iteration limit as an indicator of system instability was not a desirable methodology.

An alternative methodology is to examine the Jacobian matrix for zero-crossings. A solution proposed by [26] is to watch the Jacobian elements for a sign change which would indicate the stability limit has been reached. While this might be true theoretically, the N-R algorithm employed by MATPOWER does not produce Jacobian element zero-crossings when the system enters the point of collapse. Upon examination of the Jacobian elements of the 39-bus system as the load-factor was increased, it was determined that there were three points at which a sign change was observed in the Jacobian, none of which appear to occur when voltage collapse is observed by inspection. Two points at LF = 1.0130 and 1.2140 occur before the point of collapse and one (at LF = 1.234) occurs shortly after.

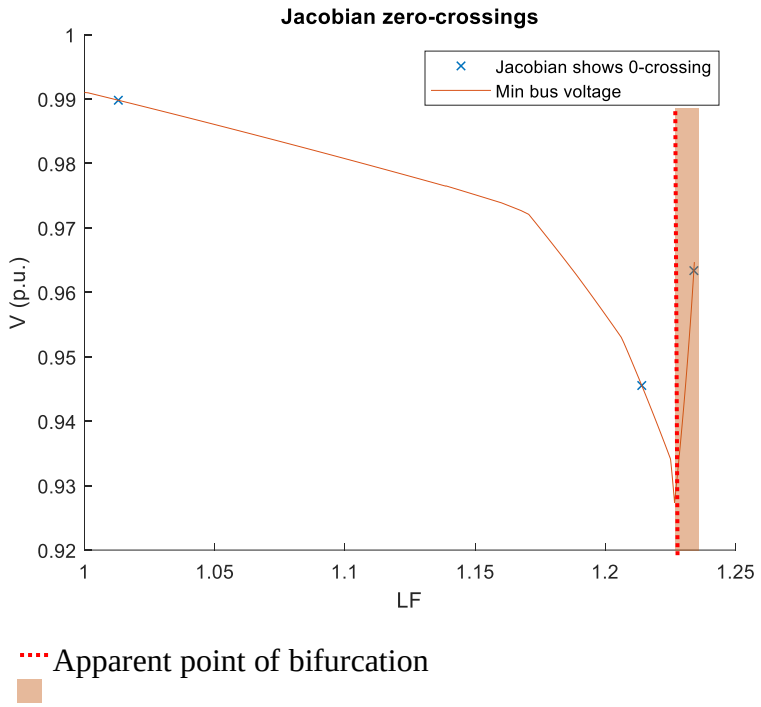


Figure 13: Minimum bus voltage and corresponding Jacobian zero-crossings

2.3.5 Jacobian Eigenvalue Analysis

A similar approach is to analyze the eigenvalues of the Jacobian matrix to determine the critical loading point. This method [48] involves tracking the minimum real eigenvalues of the Jacobian matrix as the system loading is increased. This was found to be a reliable method to determine the critical loading point that corresponds to the saddle-node bifurcation point. Figure 14 shows the behavior or the minimum eigenvalues in relation to the observed voltage and the load and Thevenin impedances for the 39-bus and 118-bus case. In both of these systems, the first eigenvalue zero-crossing (dotted vertical line) corresponds to the visual indicator of saddle-node bifurcation in the voltage graph and the point at which the trajectories of the load impedance (calculated directly) and the Thevenin impedance (estimated as described above) intersect. This

point occurs before the simulation actually fails to solve, and we can deduce that the continued solutions of the N-R power flow are mathematical artifacts rather than physical solutions.

Case 39 Case 118

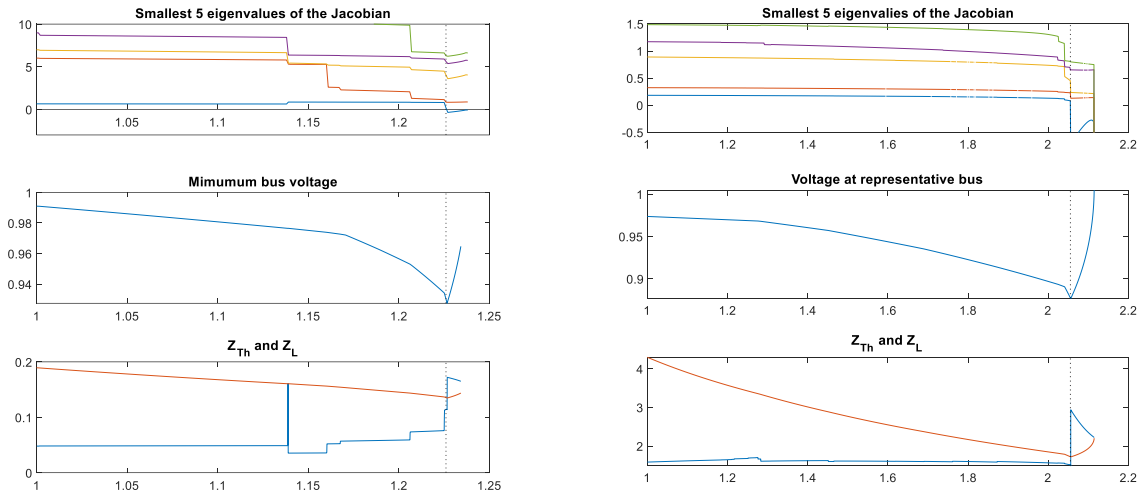


Figure 14: The zero-crossing of the reduced Jacobian minimum eigenvalue indicates the point of voltage collapse, not the convergence failure point.

The drawback of this method, at least in its implementation in this study, is that it is extremely time consuming when scaled to larger systems. In the according the MATPOWER documentation, when the Jacobian is requested, it must be reconstructed from scratch every time. Additionally, the MATLAB sparse matrix eigenvector solver (eigs) has difficulty converging with large matrices adding additional time to the computation. Ultimately, it the computationally intense nature of the implementation of this method becomes prohibitively time consuming when it must be iterated many thousands of times. Table 2 shows the observed time for the critical load factor determination. As will be discussed in the next section, the real-margin calculation requires that the critical load factor be determined calculated every time a generator reaches a

reactive limit. In the 1354-bus case, there are 260 generators, therefore we can extrapolate that using this method the true-margin calculation might take two weeks or more to complete!

2.3.6 Comparison

Tables Table 1 and Table 2 show the critical loading factor that each method described above found and the time consumed to reach this solution. Based on these results, the CPF algorithm was chosen as the most efficient and effective at finding the critical loading factor. Visually, it matched well with the bifurcation point, without carrying the risk of over-estimating the margin that the N-R method carries in some cases. It was also by far the fastest algorithm for finding the critical loading which is important because the algorithm must be iterated many times in order to account for generator reactive limits.

Table 1: Critical load factors determined by each method

	Case 9	Case 39	Case 118	Case 1354
NR 10 iteration limit	2.4307	1.3563	2.1147	1.1843
NR 20 iteration limit	2.4307	1.3563	2.1147	1.1883
First eigenvalue crosses zero	2.4305	1.2265	2.056	1.166
CPF with PQ limits enforced	2.430472	1.2342	2.1145	1.1842

Since there was little control over the computing environment (the simulations were performed on a shared server), the values in Table 2 should be interpreted as relative measures of computational efficiency. Regardless, the CPF algorithm is the clear winner being orders of magnitude faster than the next fastest algorithm.

Table 2: Time consumed to determine the critical load factor

	Case 9	Case 39	Case 118	Case 1354
NR 10 iteration limit	26.8 s	10.3 s	39.2 s	59.4 s
NR 20 iteration limit	26.9 s	10.1 s	39.1 s	61.2 s
Smallest eigenvalue passes 0	42.4 s	447.3 s	446.0 s	5059.3 s*
CPF with PQ limits enforced	0.14 s	0.17 s	0.77 s	10.15 s

*Value estimated based on a load-factor step size of 0.001 instead of 0.0001

2.4 True Margin Determination

In order to have a methodology to compare various margin estimation methods, it's important to establish a baseline or reference margin that can be used as an evaluation criterion. In the above charts the “Actual margin” is shown to evaluate the performance of the margin estimates. This section outlines the methodologies tested and employed to calculate the critical load “Actual” or “True” margin.

It might seem a straightforward affair to determine the true system margin, since after all we have discussed at length how the maximum system loading can be determined. Why not draw a straight line on the margin vs. load-factor chart from the known point of collapse of the y-axis to the known critical loading factor on the x-axis? That is exactly how the chart would look if no generator reactive limits were enforced. It can be seen in the margin estimation charts (Figure

9:Power margin estimates on the 39-bus study case, Figure 16: True margin algorithm output) in this report that the margin predictions exhibit a step-like behavior. This is caused when generator reactive limits are reached and the generator can no longer control voltage. As we might expect, the system voltage response to increased load changes and, consequently, the estimated Thevenin equivalent impedances will be different after the PV-PQ transitions. Since, in this analysis, we are not expecting the margin estimation to be able to predict when the and where the next generator reactive limits will be reached, it would not be reasonable to judge the measurements to be in error based on future system conditions. Therefore, the algorithm which is employed to determine the "true" margin should only consider system conditions as they exist at the system operating point being examined. This iterative algorithm to accomplish this is shown in the flow chart in Figure 15. The premise is that the system is solved as the load-factor in incremented until a generator reactive limit is reached, indicating a PV-PQ transition. The reactive limits for all PV buses are disabled and the critical load factor is determined using one of the methods described in the previous section.

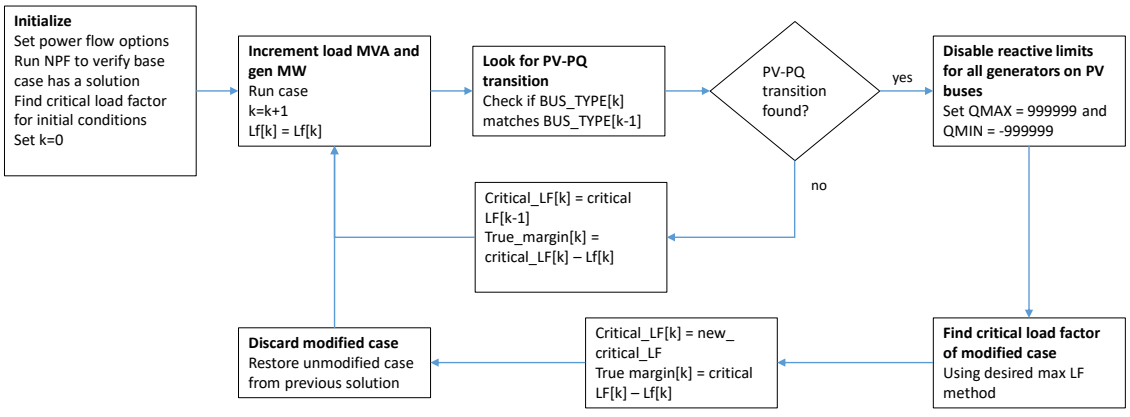


Figure 15: Algorithm for determining true margin

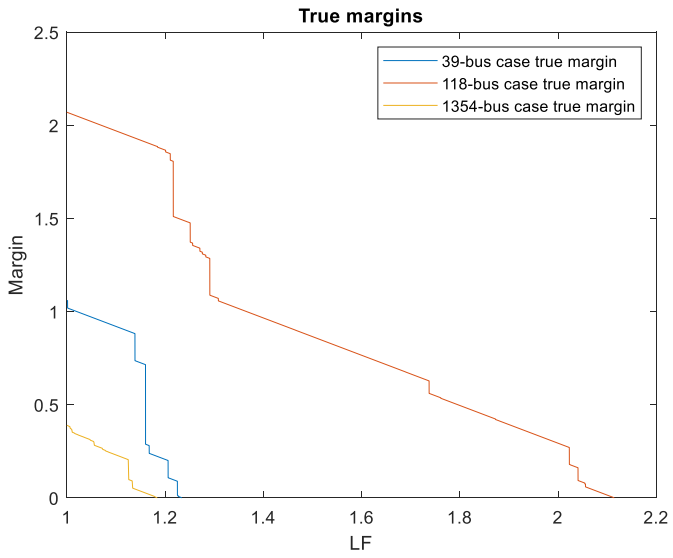


Figure 16: True margin algorithm output

Figure 16 above shows the result of the true margin calculation algorithm for the 39-bus, 118-bus and 1354 bus cases. As expected, they show the stair-step behavior due to the discontinuous change in system strength when a generator reactive limit is reached.

2.5 Sensor Fusion Modeling

This section describes the methods by which data fusion parameters were chosen in order to implement the Dempster-Shafer Evidential Theory model into a power systems framework. Appendix A contains a sample fusion calculation for a small network.

2.5.1 Probability Mass assignments (BPA)

For any location in the power system that is performing sensor fusion, there must be some way to ascribe a probability mass to the measurements of the nearby measurements, as well as to its own measurement. In our case, an electrical bus which is measuring the stability margin needs a way to assign probability to the measurements at surrounding buses. Once it has made a

determination about the probability that the measurements are correct, it can fuse the calculated or collected set of probability masses using Dempster's rule in order to decide which measurement (or sensor) to believe.

2.5.1.1 Consensus Algorithm

A simple method to assign probability masses, $m(S)$, to a set of sensors, S, that requires no a priori knowledge of the system is to look at the degree of agreement of one sensor to other sensors in the system. In this analysis, an algorithm is implemented at each node that compares its own measurements to that of its neighbors. An obvious downside to any such method is that it does not use any independent criteria to make assessments about the validity of its observations. If a majority of surrounding measurements are inaccurate, the algorithm will assign the highest measure of belief to an inaccurate measurement. However, it can serve to eliminate outliers and remove noise from the result set. If the most accurate measurement happens to have a large disagreement with the majority of sensors and appears to be an outlier, its results will tend to be ignored in such an algorithm. Therefore, this algorithm can only be said to be useful if we know based on prior experimental results that the best margin estimate tends not to be an outlier.

For the purposes of this study, I explored two different methods for comparing a measurement to its surrounding measurements. The first uses a simple cutoff to determine agreement. If we consider an electrical bus, b, and its set of nearby buses with sensors, S, then the margin measurements, M, can be said to have agreement when

$$\left|\frac{M(b) - M(S_j)}{M(b)}\right| \geq x\,,$$

where $\{x \in \mathbb{R} | 0 < x < 1\}$ is an arbitrary threshold for agreement. For example, if $x = 0.1$ then the measurements, $M(b)$ and $M(S_j)$, will be said to agree if the neighboring bus measurement is within 10% of the node b. If the measurements are found to disagree, then the bus b will assign a probability mass of $m = 0$ to the measurement at S_j.

If the buses' measurements agree, the probability mass of the measurement at S_j is normalized based on the total number of measurement buses in the system so that the total assigned probability masses can never exceed 1, even if the set of measurements includes all of the measurement nodes in the system. If the measurement of node b agrees with its neighbor, S_j, then it assigns a probability mass of $1/N$ where N is the total number of buses in the system.

The second method uses an exponential mapping of the disagreement factor between the measurement and its neighbors' measurements to assign probability masses. The mass assignment is calculated by the equation:

$$m(S) = \frac{e^{-|\kappa(M(b)-M(S_j))/M(b)|}}{N}.$$

The constant κ sets the rate at which the probability function decays to 0. Since the argument of the exponent is always negative, the value of the numerator will always be between 0 and 1. If the buses' measurements are in complete agreement, then the result is equivalent to the threshold method since $e^0 = 1$.

2.5.1.1.1 Applications to D-S theory

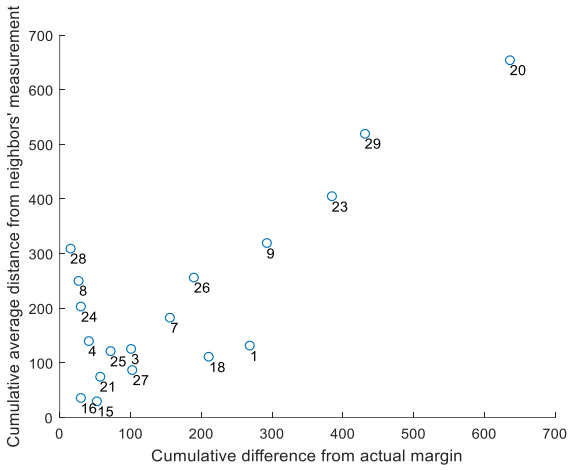


Figure 17: Relationship between disagreement and prediction error: 39-bus test case

Figure 17 illustrates the potential of employing a method to remove or discount the predictions of outliers from the data set. In this chart, the sum of the disagreement between each bus and its neighbors was summed over each iteration in the simulation and compared to the sum of the error in predicting the margin over the whole simulation. The chart provides evidence for a correlation between measurement locations with high cumulative disagreement. The correlation appears to fall apart for measurement locations with less disagreement, therefore the selection of a threshold or κ should appropriate to the system.

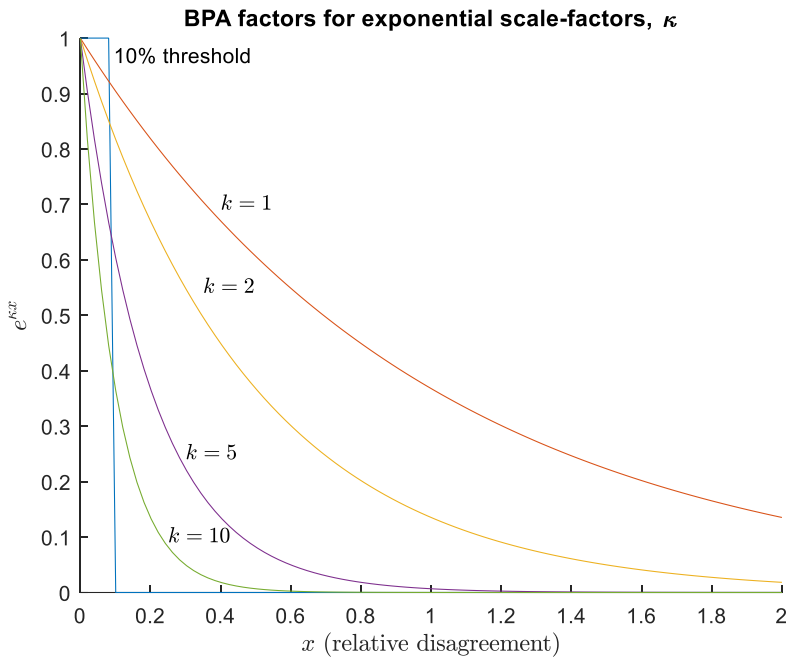


Figure 18: BPA distribution vs disagreement

2.5.1.2 Eigenvector Centrality

An additional method of assigning the probability mass to a measurement at a node looks at the degree of connectivity or "centrality" of a node as a proxy for trusting the accuracy of the measurement. Eigenvector centrality (EC) is a way of calculating the degree to which a node in a system is connected to other nodes in the system. EC was originally developed as a technique to gauge the popularity of individuals in social networks [49]. In [49], Bonacich describes a process by which each person's contribution to the popularity of others is weighted by their own popularity.

An adjacency matrix is constructed by first numbering the nodes of a network. If there are N nodes, then the matrix will be an N × N matrix. For every element A_{ij} if node i is connected to node j, then A_{ij} = 1. If the connections are bidirectional, as they are in our case, the adjacency

matrix shall be symmetric, that is, $A_{ij} = A_{ji}$. The diagonal elements A_{ii} are 0 since there is no need to define a node's connection to itself.

Let A be the adjacency matrix that defines a node network, then the centrality of unit i is given by the eigenvector function,

$$\lambda e_i = \sum_j A_{ij} e_j$$

Or in matrix notation,

$$\lambda e = Ae$$

where e is the eigenvector and λ is the eigenvalue.

The centrality of a node is defined as the sum of each measure of centrality connected to adjacent nodes. This sum would tend towards infinity as the solution iterates. In order to avoid this, for each iteration, the centrality measure is normalized by the maximum eigenvalue. The centrality, C_e, can be then calculated by iterating the following equation beginning with an arbitrary column vector (usually vector of ones):

$$C_e = e_i = \frac{1}{\lambda_{max}} \sum_j A_{ij} e_j$$

For a power flow case, the matrix A can be constructed from the nodal admittance matrix (Y_{bus}). The Y_{bus} matrix contains information about the admittance, y_{ij}, (inverse of the impedance)

between each node, i and j, in a power system. If there is no connection between the nodes, then the corresponding entry in the matrix is 0. The Y_{bus} matrix is given by

$$Y_{ij} = \begin{cases} y_i + \sum_{\forall k | k \neq i} y_{ik} & \text{if } i = j \\ -y_{ij} & \text{if } i \neq j \end{cases}$$

Since the Y_{bus} matrix is readily available in many power flow software applications, it is a convenient starting point to construct the adjacency matrix. This is done by setting the non-zero matrix elements equal to 1 and the diagonal elements to 0. To express this in equation form, we can take advantage of the sign function which returns a 1 or a -1 for a non-zero input and 0 when the input is 0. The admittance between nodes should be positive for any realistic power flow case, but the elementwise absolute value is taken as well for good measure:

$$A = \text{sgn}\left|\text{Im}\left(Y_{bus} - diag(Y_{bus})\right)\right|$$

Where $\text{sgn}(y)$ is the sign function defined as

$$\text{sgn}(x) = \begin{cases} -1 : x < 0 \\ 0 : x \leq 0 \\ 1 : x > 0 \end{cases}$$

and $\mathrm{diag}(x)$ is a function[3] that returns the matrix with only its diagonal elements, setting all other elements to zero. $\mathrm{Im}(x)$ returns the imaginary part of the complex number such that $\mathrm{Im}(a + jb) = b$.

When applying the eigenvector centrality algorithm to analyzing power systems, we can take the theory one step further since not all connections in power systems are equal. If we posit that a stronger connection between two nodes means that they exert more influence over one another, then the strength of the connection should be considered when calculating centrality. Consider a system of nodes and connections described by the adjacency matrix A. Now imagine creating and additional connection between two already connected nodes i and j. Instead of $A_{ij} = A_{ji} = 1$, we would have $A_{ij} = A_{ji} = 2$. Alternatively, $A_{ij} = A_{ji} = 1$ might retain its value while all other connections would be represented by ½. Since the connection strength is a relative measure, either of these would give the same result after applying the EC algorithm. If we were describing an electrical system of identical conductors, we would observe that the impedance of the double connection has been halved, or equivalently, that the admittance has been doubled. At this point, it is probably obvious to the reader that the Y_{bus} matrix of a power flow case, which is a matrix of the admittance between each connected node, contains precisely this information about relative strength of connections. Indeed, researchers Wang, et al [50] reach this exact same conclusion of how EC can relate to power systems. Since admittance is expressed as complex number and the reactance of lines dominates the impedance of transmission lines, the imaginary

[3] In MATLAB the "diag" function serves a dual purpose. When the argument is a vector of length N, it returns an N×N matrix with the values of the vector along the matrix diagonal. When the argument is an N×N matrix, it returns a vector of length N containing the diagonal elements. Thus the function described above can be accomplished by calling "diag(diag(Ybus))".

parts of the Y_{bus} can be substituted in into the admittance matrix as a simplifying assumption for the EC calculation:

$$A_{\text{Ybus}} = \left|\text{Im}\left(Y_{bus} - diag(Y_{bus})\right)\right|$$

The resulting EC calculation can be expressed as

$$C_e = e_i = \frac{1}{\lambda_{\max}} \sum_j A_{\text{Ybus},ij} e_j \, .$$

2.5.1.2.1 Applicability to a Distributed Computational Scheme

It is worth noting that the EC algorithm translates naturally to a distributed computation framework. For the purposes of this study, it is convenient to use the modified admittance matrix or adjacency matrix when performing the EC calculation since the system is being examined and simulated as a whole. From a single node's perspective, the calculation involves summing the centrality measures of the nodes which are directly connected to it, optionally weighting the sum by the impedances of the connections between them. Therefore, a node does not require any special knowledge of system-wide topology. The only system-wide parameter needed is maximum eigenvalue for the normalization step of the algorithm. Since the centrality is determined purely by the system topology, it would change relatively slowly compared to other system parameters. Only when there is a nearby physical topology change, such as a line outage, a switching event, or a tap setting change (in the case of the Y-weighted EC calculation) would there be any change in the EC parameter. Therefore, the maximum eigenvalue could be updated fairly infrequently. Examples of such schemes could be by point-to-point relaying across the

system or perhaps hourly updates from a centralized server. Or the node performing the EC calculation could request on-demand updates when it becomes aware of a topology change.

Additionally, it would be impractical to install a device that calculates EC at every single system node. This can be solved by performing EC calculations at various distributed nodes throughout the system which are responsible for tracking a subset of nearby nodes and the connectivity between them.

2.5.1.2.2 Centrality Algorithm Output Example

In this section the EC algorithms described above are applied to the IEEE 14-bus test case and the results are discussed. The test case consists of high- and medium-voltage levels. High-voltage in this context refers to 345-kV buses, and medium-voltage refers to anything above 60 kV but below 345-kV. In this case, MV buses are 138-kV buses.

Table 3: IEEE 14 bus test case eigenvector centrality algorithm results

	Unweighted		**Ybus Weighted**	
Bus	Centrality	Rank	Centrality	Rank
1 (HV)	0.5034	7	0.5311	4
2 (HV)	0.8584	2	0.7039	3
3 (HV)	0.5515	5	0.2991	5
4 (HV)	1.0000	1	0.9984	2
5 (HV)	0.8373	3	1.0000	1
6 (MV)	0.4591	8	0.1670	8
7 (MV)	0.5209	6	0.2377	6
8 (MV)	0.1546	14	0.0478	10
9 (MV)	0.6004	4	0.1696	7
10 (MV)	0.2401	11	0.0678	9
11 (MV)	0.2077	13	0.0349	12
12 (MV)	0.2197	12	0.0220	13
13 (MV)	0.2795	9	0.0397	11
14 (MV)	0.2615	10	0.0215	14

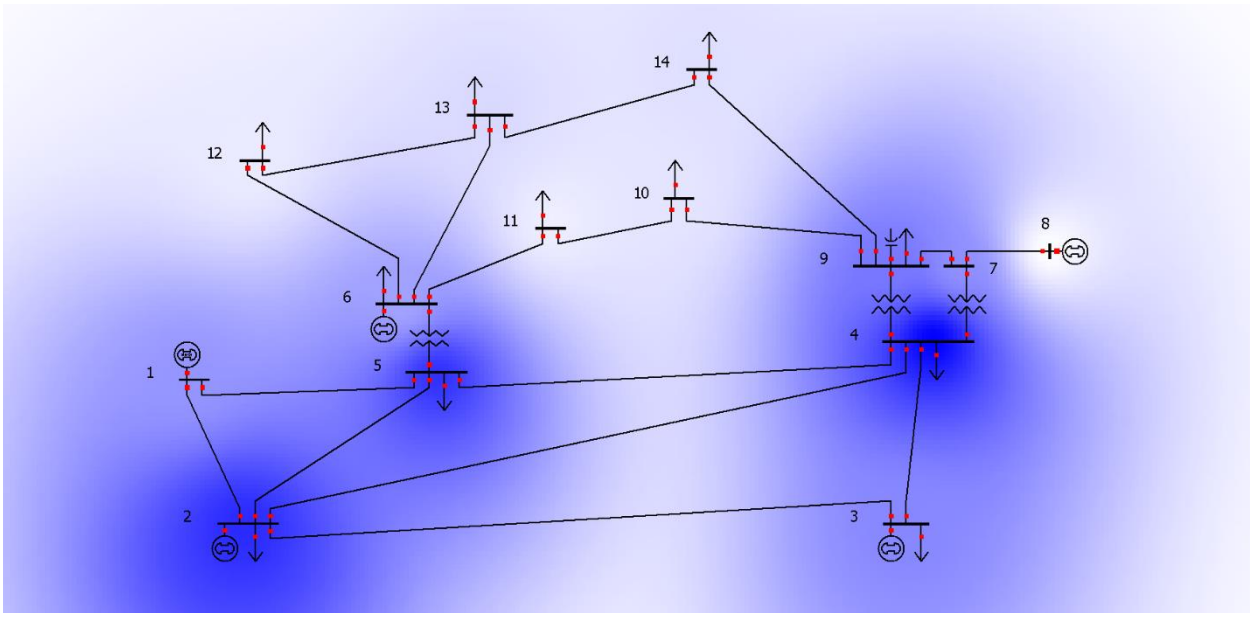


Figure 19: Unweighted Centrality measure of 14-bus system contour plot

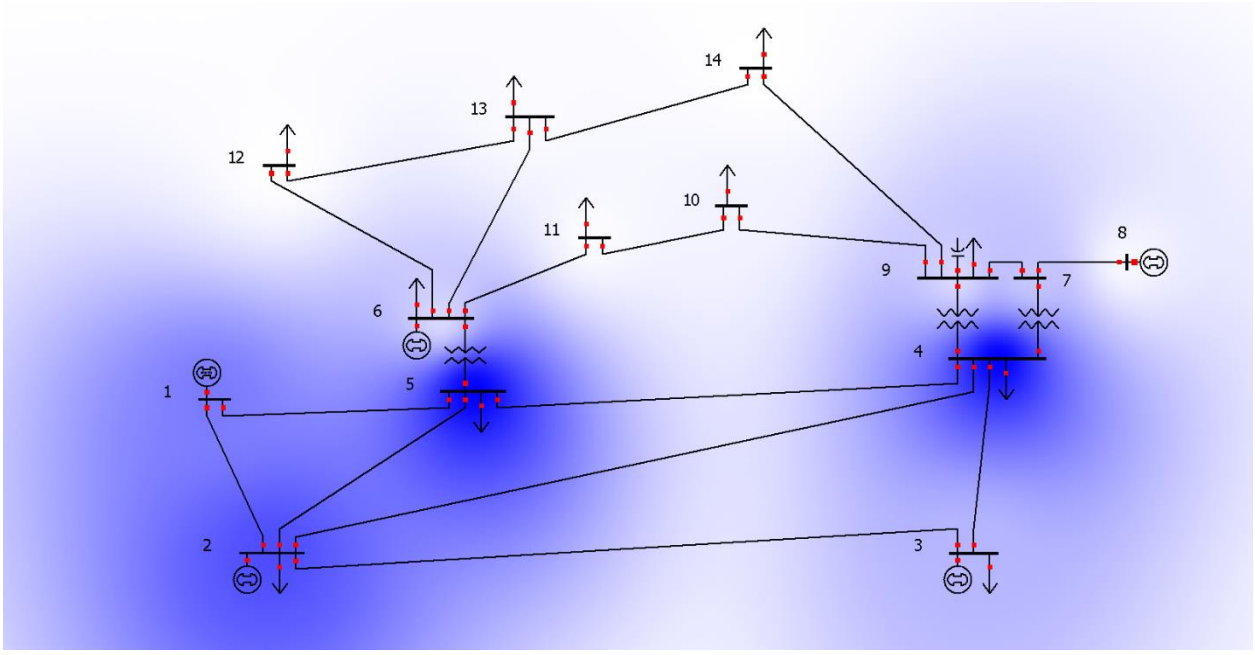


Figure 20: Ybus weighted centrality measure of 14-bus system contour plot

We observe from the results of this small test case that, buses which are ranked highly central in the unweighted centrality calculation tend to also be highly ranked in the Y-weighted calculation. As we would might expect, in the weighted calculation, all of the top 5 buses are the 5 high-voltage buses In the unweighted calculation, four of the five highest central buses are HV buses. The preference given to high voltage buses reflects to the topology of power systems; HV buses tend to be more centrally connected and have more connections to them as high-voltage parts of an interconnection form the backbone of the grid. In the case of the weighted calculation, the preference is reinforced since higher voltage parts of the system will tend to have larger transmission lines and lower impedance values, although they may also tend to be longer which would result in higher impedance values.

Table 4: Top ten centrality values, 39 bus case

Bus	Unweighted	Bus	Weighted
16	1	6	1
4	0.71866	5	0.938183
17	0.717655	7	0.41546
26	0.679329	11	0.414372
6	0.674446	8	0.379972
5	0.671674	10	0.294881
14	0.669835	4	0.172684
3	0.654868	13	0.163916
2	0.609579	31	0.08165
15	0.600269	14	0.065072

Table 5:Top ten centrality values 118 bus case

Bus	Unweighted	Bus	Weighted
49	1	68	1
69	0.924705	116	0.945578
77	0.710505	65	0.246948
54	0.656763	81	0.191977
75	0.649447	**69**	0.111974
80	0.587372	64	0.033246
100	0.586579	66	0.027819
59	0.583738	**80**	0.022241
47	0.539054	38	0.009749
70	0.530949	**77**	0.006938

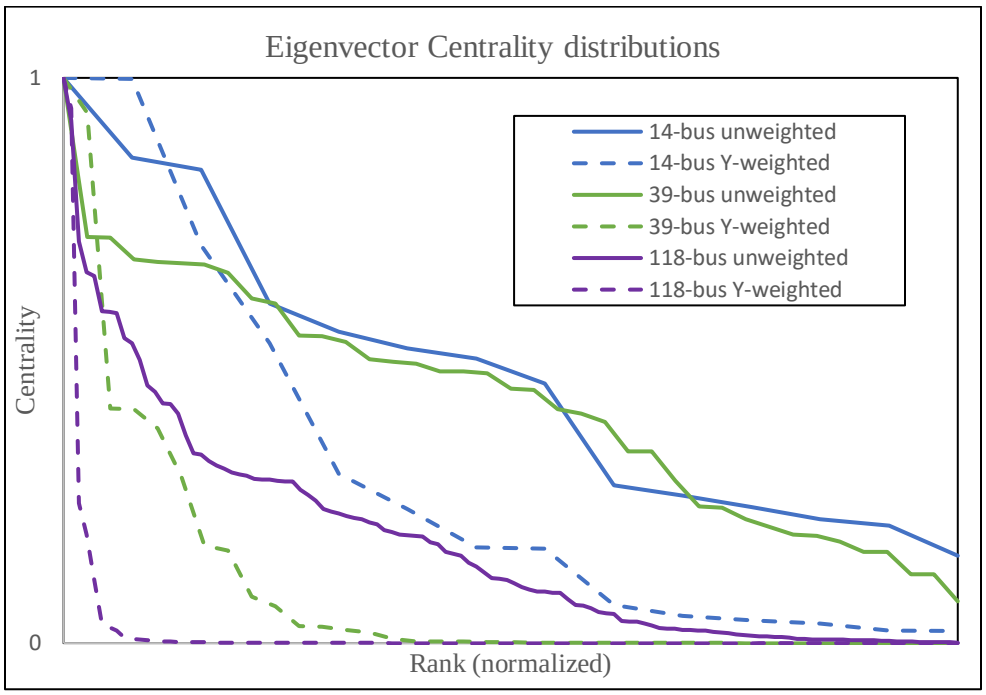


Figure 21: Eigenvector centrality distribution. For larger systems, the Y-weighted centrality measure falls off extremely rapidly.

2.5.1.2.3 Usefulness for D-S Combination

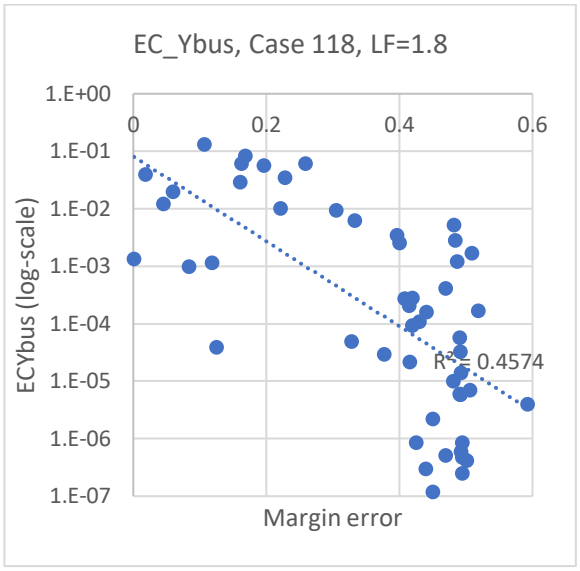


Figure 22: Margin error vs Y-weighted centrality

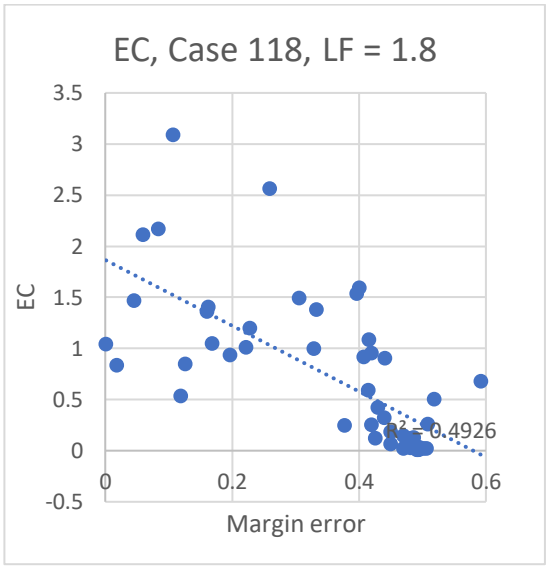


Figure 23: Margin error vs unweighted centrality

The purpose of applying this algorithm is in hopes that it can be used to help identify buses with better margin estimates. The assumption being tested in this case is that buses that are more central to the system will have a better measure of the stability of the system than a remote, less central bus. To this end, if it can be observed that there is a positive correlation between the eigenvector centrality (EC) assignment and the bus accuracy, then it might be a useful parameter in assigning a probability mass, or BPA. As a cursory check on whether or not the variable has potential to be useful in BPA determination the centrality is plotted against the prediction error for each bus in the 118-bus case at an arbitrary loading factor in Figure 22 an Figure 23. Since the centrality assignments decay extremely rapidly when using the modified Ybus matrix in place of the adjacency matrix, the y-axis is plotted on a logarithmic scale. There appears to be a

negative correlation between the centrality and the error, which suggests that a higher centrality assignment would tend to correspond to a better margin prediction at a VIP bus.

In this study EC was incorporated into the BPA as a simple weight. The way in which the range is defined, $\{EC \in x \mid 0 < x \leq 1\}$, means that it requires no additional normalization to be applied to the BPAs. The range of BPAs is defined as $\{BPA \in \theta \mid 0 \leq \theta < 1/N\}$ where N is the cardinality of the set of BPAs so that the sum of all BPAs in a set cannot exceed unity. With the range of the EC variable as defined above, this requirement is preserved: $\{EC \times BPA \in y \mid 0 \leq y \leq 1/N\}$.

The Y-weighted EC calculation was not used in subsequent simulations. Although, on its face, it would seem to be a more physically representative value of the system configuration, in practice (as can be observed in Figure 21) the EC assignment decays too quickly with the rank of the EC assignment maximum value to be practically applied as a weight to for BPA calculation. For larger systems the rate of decay is even faster. For example, in the 1354-bus system studied in this report, the median Y-bus weighted EC value is 9.3×10^{-18} and the minimum value is 7.1×10^{-21}. In this dataset, only the top 2.5% of buses have a Y-weighted EC value greater than the *minimum* unweighted EC value observed (9.5×10^{-8}). Thus, if the Y-weighed EC was applied, it would effectively be multiplying most BPAs by approximately 0.

2.5.2 Simulation Results

In this section the simulation results that were generated by employing the models and method described in this text. The simulations were run with variations in model parameters to produce fused datasets which were compared to the predictive performance of the VIP prediction results that has not undergone any data fusion. The main criteria for evaluation of the performance of

the data fusion algorithm is the prediction accuracy at each bus post-fusion as compared to pre-fusion. To this end, the prediction results reported by each VIP bus are averaged over each system load-factor step. The averages are plotted in the charts below. Both average absolute error and average per-unit error are presented in the charts below. While the relative rank of the performance of each dataset does not change, both methods are useful in understanding the performance. Generally, a per-unit or percent error would be preferred, however the per-unit error tends to blow up as the load-factor approaches the critical loading since the denominator of the percent error calculation is the "true" margin which approaches zero at this operating point.

In addition to the plots the total average per-unit errors have been calculated for evaluating the performance of the dataset and the total average per-unit error of the un-fused data is evaluated as well.

2.5.2.1 39-Bus Test Case

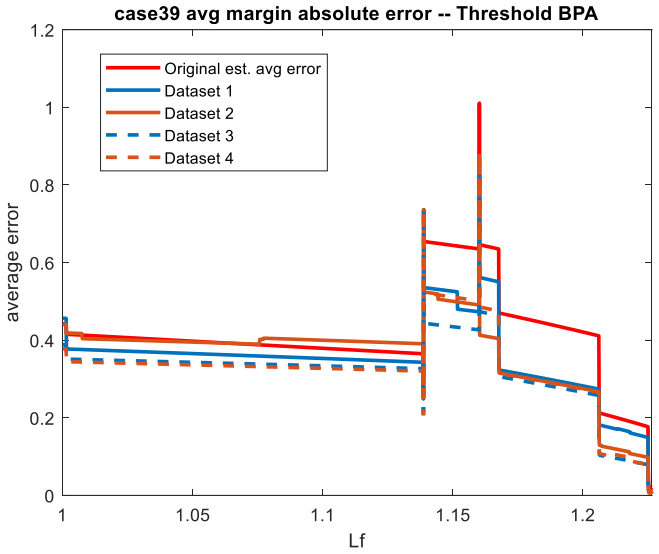


Figure 24: Average absolute error for 39-bus simulation datasets using the threshold BPA method

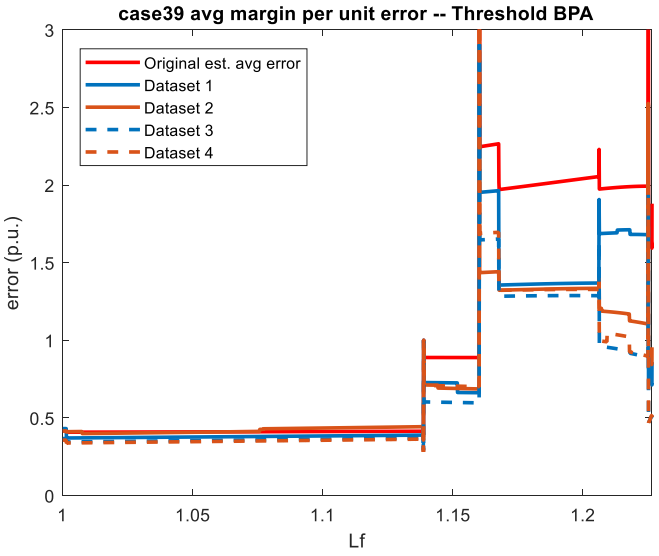


Figure 25: Average per-unit error for 39-bus simulation datasets using the threshold BPA method

Datasets:

Unfused avg. per unit error:

0.932103

No.	T	Weight	Avg tot err
1	0.1	UW	0.743278
2	0.2	UW	0.700525
3	0.1	CW	0.631375
4	0.2	CW	0.649171

Key
T – Agreement threshold
UW – unweighted BPA
CW – centrality weighted BPA

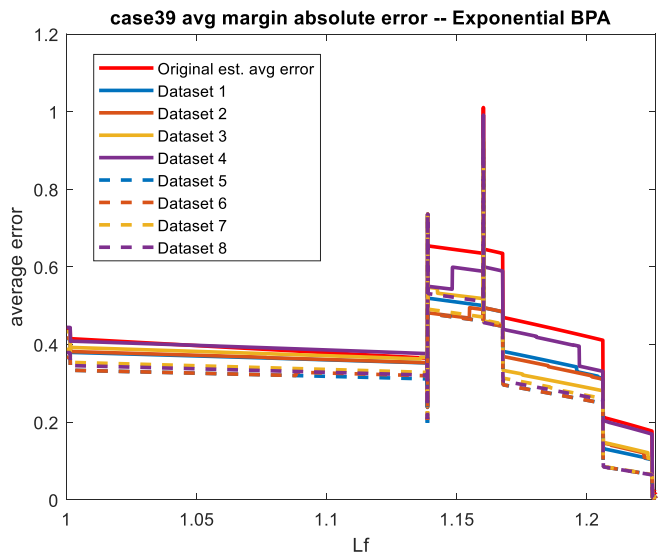


Figure 26: Average absolute error for 39-bus simulation datasets using the exponential BPA method

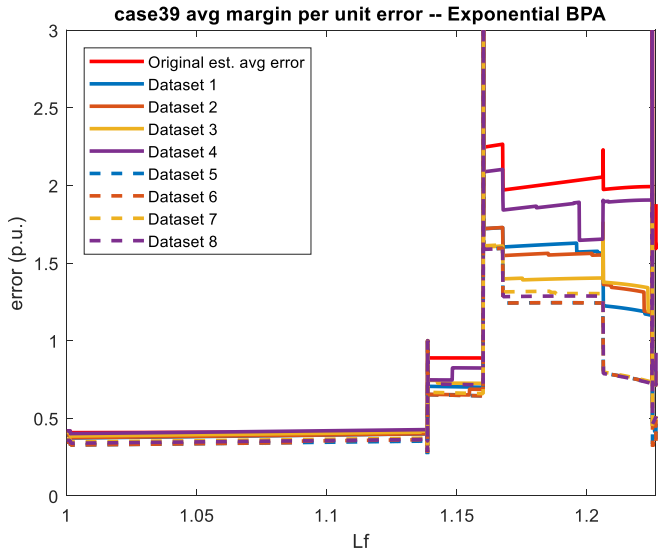


Figure 27: Average per-unit error for 39-bus simulation datasets using the exponential BPA method

Datasets:

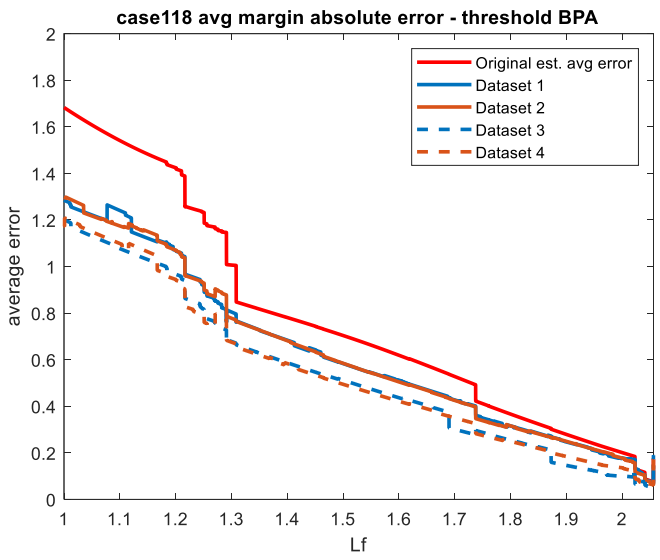
Unfused average per unit error:

0.932103

No.	κ	Weight	Avg tot err
1	1	UW	0.737657
2	2	UW	0.735485
3	5	UW	0.722999
4	10	UW	0.871838
5	1	CW	0.601064
6	2	CW	0.603016
7	5	CW	0.627928
8	10	CW	0.62282

Key
κ – exponential scale factor
UW – unweighted
CW - centrality weighted

2.5.2.2 118-Bus Test Case



Datasets:

Unfused avg. per unit error:

0.784232

No.	T	Weight	Avg. Err
1	0.1	UW	0.650873
2	0.2	UW	0.648125
3	0.1	CW	0.529967
4	0.2	CW	0.542833

Key
T – Agreement threshold
UW – unweighted BPA
CW – centrality weighted BPA

Figure 28: Average absolute error for 118-bus simulation datasets using the threshold BPA method

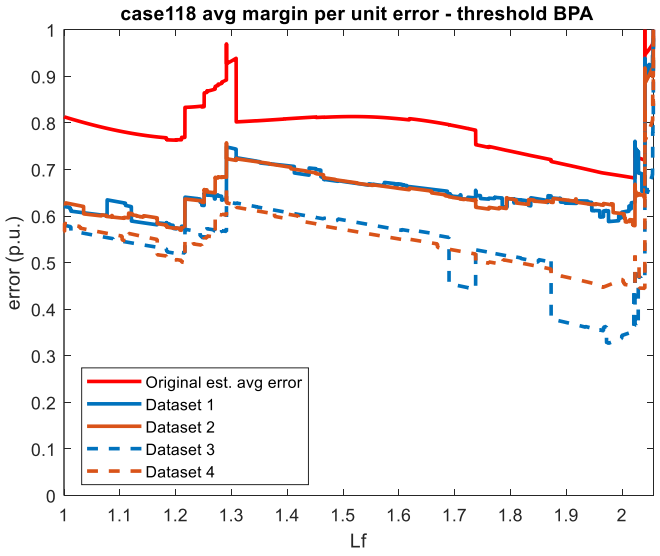


Figure 29: Average per-unit error for 118-bus simulation datasets using the threshold BPA method

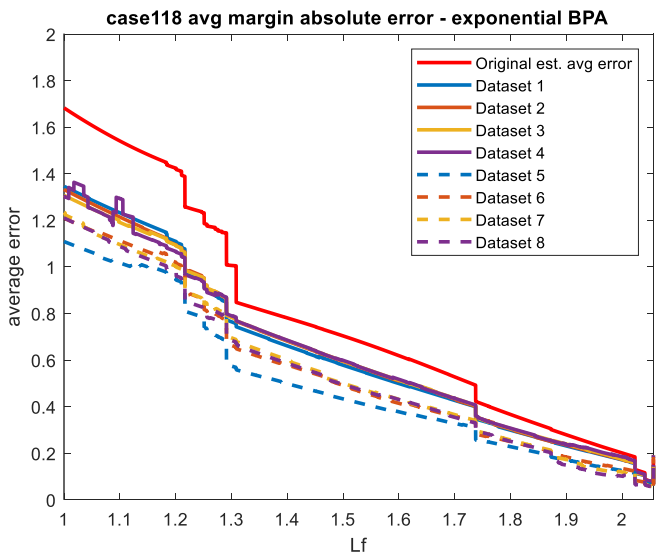


Figure 30: Average absolute error for 118-bus simulation datasets using the exponential BPA method

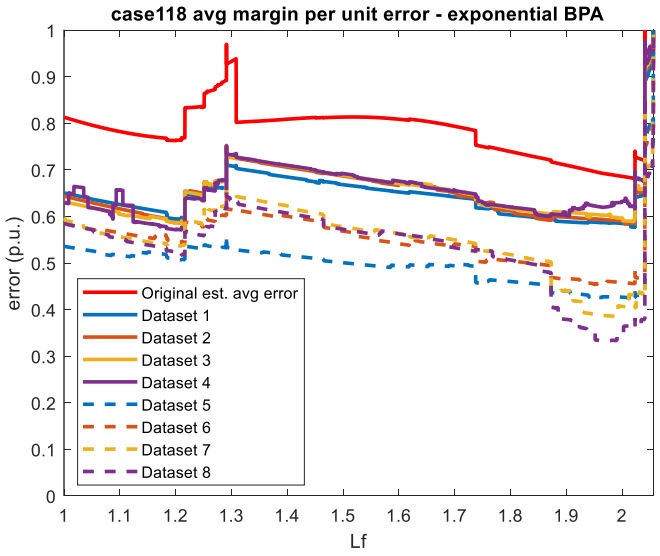


Figure 31: Average relative error for 118-bus simulation datasets using the exponential BPA method

Datasets:

Unfused average per unit error: 0.784232

No.	κ	Weight	Avg Err.
1	1	UW	0.641734
2	2	UW	0.649887
3	5	UW	0.651438
4	10	UW	0.654892
5	1	CW	0.496938
6	2	CW	0.547722
7	5	CW	0.548194
8	10	CW	0.531628

Key
κ – exponential scale factor
UW – unweighted
CW - centrality weighted

2.5.2.3 1354-Bus Test Case

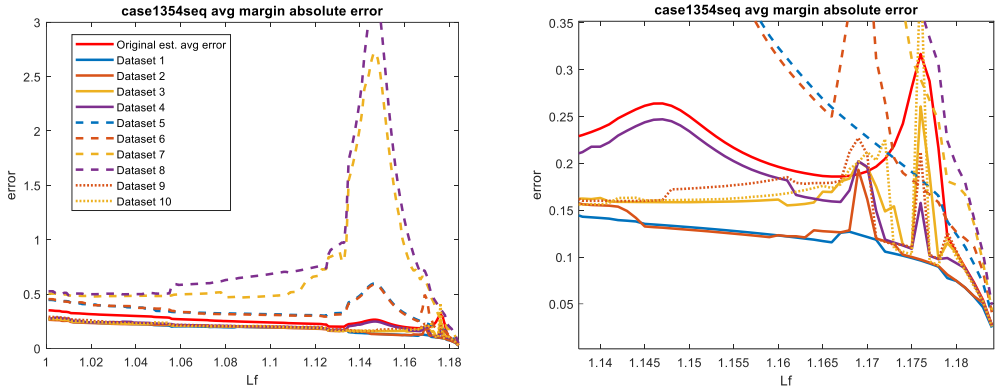


Figure 32: Average absolute error for fusion methods with 1354-bus test case (left) and detail near the critical loading (right)

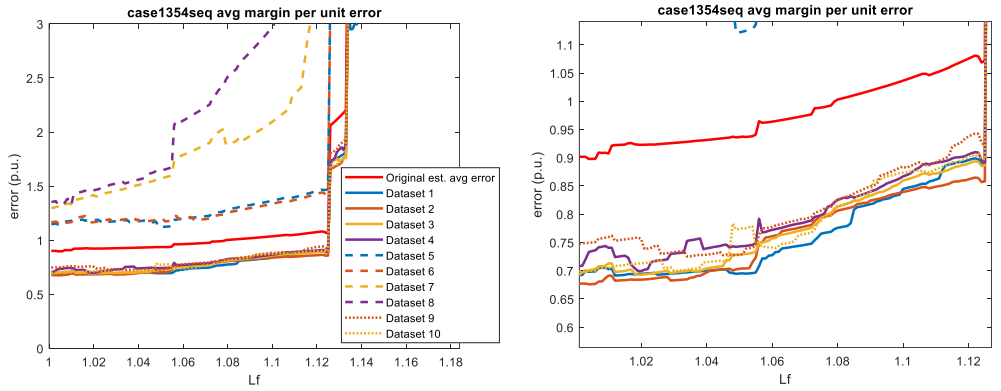


Figure 33: Average per-unit error for fusion methods with 1354-bus test case (left) and detail near the critical loading (right)

Table 6: Dataset descriptions for Figure 32 and Figure 33

Dataset	Method	parameters	Average per unit error
Unfused	Simple average		2.20798
1	Exponential BPA, unweighted	$\kappa = 1$	1.47211
2	Exponential BPA, unweighted	$\kappa = 2$	1.48698
3	Exponential BPA, unweighted	$\kappa = 5$	1.68066
4	Exponential BPA, unweighted	$\kappa = 10$	1.93657
5	Exponential BPA, centrality-weighted	$\kappa = 1$	3.67127
6	Exponential BPA, centrality-weighted	$\kappa = 2$	3.61608
7	Exponential BPA, centrality-weighted	$\kappa = 5$	11.3815
8	Exponential BPA, centrality-weighted	$\kappa = 10$	13.4548
9	Threshold BPA, unweighted	$T = 0.1$	1.78412
10	Threshold BPA, unweighted	$T = 0.2$	1.70768

3 CONCLUSIONS

Overall, sensor fusion techniques were found to be able to reduce the prediction error in the systems studied. Table 7 summarizes the best data fusion runs from the set above. Based on the observed simulations, the exponential BPA calculation for the consensus algorithm with a scale-factor (κ) of 1 was the best performing approach. Across the board, the consensus algorithm on its was able to reduce the prediction error in every data fusion run that was performed. In each case, data fusion was able to reduce the prediction error by between 33.3% and 36.6%. While the practical value of reducing prediction error from, say, 220% to 147% (as in the 1354-bus case) might be debatable, it is evident that the data fusion consensus algorithm was able to consistently reduce prediction error.

The addition of the eigenvector-centrality weighting was not consistently able improve prediction performance. While it did incrementally improve the performance in the 39- and 118-bus cases, it caused the error to increase in the 1354-bus case. The reason for this is reversal unclear, but it implies that the better-connected buses in the 1354-bus case are not more reliable than other buses. One possibility is that in the 1354-bus case, generators tend to be more centrally connected thus reducing voltage variations of nearby simulated VIPs, leading to more inaccurate Thevenin parameters being favored.

Table 7: Best performing models for each case

Case	Unfused average error	Best observed average error after fusion	Percent improvement	Method/Parameters with best result
39-bus	0.932103	0.601064	35.5%	Exponential BPA, centrality weighted, $\kappa = 1$
118-bus	0.784232	0.496938	36.6%	Exponential BPA, centrality weighted, $\kappa = 1$
1354-bus	2.20798	1.47211	33.3%	Exponential BPA, unweighted, $\kappa = 1$

3.1 Opportunities for Future Study

While the results of this study are far from groundbreaking, I believe the data fusion approaches explored in this work present an interesting and useful framework upon which other studies could be based. The main hinderance here is that a reliable method for estimating a buses' margin prediction accuracy based on a topological or electrical parameter was never identified. In fact, a great deal of time in the research that preceded this study was spent searching a parameter or combinations of parameters that would have some predictive accuracy. As the number of buses in a network increases, the data becomes increasingly difficult to analyze. The nature of this problem makes the methods explored here an ideal candidate for a machine-learning framework to identify predictive parameters since machine learning tends to excel at solving "big data" problems.

In this study's algorithm to determine confidence measures, the fused BPAs are discarded every iteration and recalculated from scratch. This was approach was chosen so that the VIPs would respond instantaneously to a change in system conditions. However, each measurement node could remember the prior results of the data fusion and feed this into the next calculation. Using this approach, the system would eventually converge to a single consensus value. In principle, if an accurate correct predictive method is found, the system could converge to a single correct value. However, in a dynamic system, the success of this approach would depend on how quickly the correct prediction would be able to propagate through the system.

Another opportunity for improvement could be to allow margin estimates to be a range of values. In this study, the margin predictions are treated as singular, incompatible propositions. In other words, if the estimated margin is X, it cannot be Y. Dempster-Shafer theory provides a framework for fusing compatible propositions by representing the margin estimates as range

instead of a single value. In this case if the margin estimate X is a range of values, then the margin estimate Y can also be correct or overlap with Y. One approach could be to interpret the Z-margin and the P-margin as upper and lower limits of the range of likely margin values. If it can be shown that the true margin is likely to fall within the range bordered by these two values, this could be a promising approach.

REFERENCES

[1] T. Van Cutsem and C. Vournas, *Voltage stability of electric power systems*. Springer Science & Business Media, 1998.

[2] J. J. Grainger, W. D. Stevenson, and W. D. Stevenson, *Power system analysis*. 2003.

[3] A. M. Chebbo, M. R. Irving, and M. J. H. Sterling, "Voltage collapse proximity indicator: behaviour and implications," in *IEE Proceedings C-Generation, Transmission and Distribution*, 1992 1992, vol. 139, 3 ed., pp. 241-252.

[4] M. Haque, "A fast method for determining the voltage stability limit of a power system," *Electric power systems research,* vol. 32, no. 1, pp. 35-43, 1995.

[5] M. Haque, "Novel method of assessing voltage stability of a power system using stability boundary in P–Q plane," *Electric Power Systems Research,* vol. 64, no. 1, pp. 35-40, 2003.

[6] D. L. Hall and S. A. H. McMullen, *Mathematical techniques in multisensor data fusion*. Artech House, 2004.

[7] P. K. Varshney, "Multisensor data fusion," *Electronics & Communication Engineering Journal,* vol. 9, no. 6, pp. 245-253, 1997.

[8] H. Li, B. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," *Graphical models and image processing,* vol. 57, no. 3, pp. 235-245, 1995.

[9] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE,* vol. 85, no. 1, pp. 6-23, 1997.

[10] A. O. Boudraa, L. Bentabet, F. Salzenstein, and L. Guillon, "Dempster-Shafer's basic probability assignment based on fuzzy membership functions," *Electronics Letters on Computer Vision and Image Analysis,* vol. 4, no. 1, pp. 1-9, 2004.

[11] K. Sentz and S. Ferson, *Combination of evidence in Dempster-Shafer theory*. Citeseer, 2002.

[12] P. Kundar *et al.*, "Definition and classification of power system stability (vol 18, pg 1387, 2004)," *IEEE Transactions on Power Systems,* vol. 19, no. 4, pp. 2124-2124, 2004.

[13] K. T. Vu and C.-C. Liu, "Dynamic mechanisms of voltage collapse," *Systems & Control Letters,* vol. 15, no. 4, pp. 329-338, 1990/11/01/ 1990, doi: https://doi.org/10.1016/0167-6911(90)90106-5.

[14] M. Begovic, D. Novosel, and M. Milisavljevic, "Trends in power system protection and control," *Decision Support Systems,* vol. 30, no. 3, pp. 269-278, 2001.

[15] J. Adams *et al.*, *ERCOT ISO's experiences in handling voltage related issues in the control center*. 2011, pp. 1-4.

[16] K. T. Vu, D. E. Julian, J. O. Gjerde, and M. M. Saha, "Applications and methods for voltage instability predictor (VIP)," ed: Google Patents, 2001.

[17] K. T. Vu and D. Novosel, "Voltage Instability Predictor (VIP)--Method and System for Performing Adaptive Control to Improve Voltage Stability in Power Systems," USA Patent 6,219,591, 2001.

[18] K. Vu, M. Begovic, D. Novosel, and M. M. Saha, "Use of Local Measurements for Estimates Voltage-Stability Margin," *IEEE Transactions on Power Systems,* vol. 14, no. 3, pp. 1029-1035, 1999.

[19] D. Julian, R. P. Schulz, K. Vu, W. H. Quaintance, N. B. Bhatt, and D. Novosel, "Quantifying proximity to voltage collapse using the voltage instability predictor (VIP)," in *Power Engineering Society Summer Meeting, 2000. IEEE,* 2000, vol. 2: IEEE, pp. 931-936.

[20] B. Gu, J. Dominic, J.-S. Lai, C.-L. Chen, T. LaBella, and B. Chen, "High reliability and efficiency single-phase transformerless inverter for grid-connected photovoltaic systems," *Power Electronics, IEEE Transactions on,* vol. 28, no. 5, pp. 2235-2245, 2013.

[21] M. Haque, "Use of V–I characteristic as a tool to assess the static voltage stability limit of a power system," *IEE Proceedings-Generation, Transmission and Distribution,* vol. 151, no. 1, pp. 1-7, 2004.

[22] T. An, S. Zhou, J. Yu, W. Lu, and Y. Zhang, "Research on illed-conditioned equations in tracking thevenin equivalent parameters with local measurements," in *2006 International Conference on Power System Technology,* 2006 2006, pp. 1-4.

[23] K. T. Vu, D. E. Julian, J. O. Gjerde, and M. M. Saha, "Applications and Methods for Voltage Instability Predictor (VIP)," USA Patent 6,249,719, 2001.

[24] B. Milosevic and M. Begović, "Voltage-stability protection and control using a wide-area network of phasor measurements," *Power Systems, IEEE Transactions on,* vol. 18, no. 1, pp. 121-127, 2003.

[25] L. Warland and A. T. Holen, "A voltage instability predictor using local area measurements (VIP++)," in *Power Tech Proceedings, 2001 IEEE Porto*, 2001, vol. 2, pp. 6-pp.

[26] V. A. Venikov, V. A. Stroev, V. I. Idelchick, and V. I. Tarasov, "Estimation of electrical power system steady-state stability in load flow calculations," *IEEE Transactions on Power Apparatus and Systems,* vol. 94, no. 3, pp. 1034-1041, 1975, doi: 10.1109/t-pas.1975.31937.
[27] Y. Tamura, H. Mori, and S. Iwamoto, "Relationship between voltage instability and multiple load flow solutions in electric power systems," *IEEE Transactions on power apparatus and systems,* no. 5, pp. 1115-1125, 1983.
[28] I. Hano, Y. Tamura, S. Narita, and K. Matsumoto, "Real time control of system voltage and reactive power," *IEEE Transactions on Power Apparatus and Systems,* no. 10, pp. 1544-1559, 1969.
[29] P.-A. Lof, T. Smed, G. Andersson, and D. Hill, "Fast calculation of a voltage stability index," *IEEE Transactions on Power Systems,* vol. 7, no. 1, pp. 54-64, 1992.
[30] P.-A. Lof, G. Andersson, and D. Hill, "Voltage stability indices for stressed power systems," *IEEE transactions on power systems,* vol. 8, no. 1, pp. 326-335, 1993.
[31] H.-D. Chiang, I. Dobson, R. J. Thomas, J. S. Thorp, and L. Fekih-Ahmed, "On voltage collapse in electric power systems," *IEEE Transactions on Power systems,* vol. 5, no. 2, pp. 601-611, 1990.
[32] N. T. Hawkins, G. Shackshaft, and M. J. Short, "On-line algorithms for the avoidance of voltage collapse: reactive power management and voltage collapse margin assessment," in *Power System Monitoring and Control, 1991., Third International Conference on*, 1991, pp. 134-139.
[33] L. A. Klein, *Sensor and Data Fusion: a tool for information assessment and decision making*. Bellinham, Wash.: SPIE Press, 2004.
[34] J. Bernoulli and J. Bernoulli, *The art of conjecturing, together with Letter to a friend on sets in court tennis*. JHU Press, 2006.
[35] R. R. Yager and L. Liu, *Classic works of the Dempster-Shafer theory of belief functions*. Springer, 2008.
[36] G. Shafer, *A mathematical theory of evidence*. Princeton university press Princeton, 1976.
[37] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *The annals of mathematical statistics,* pp. 325-339, 1967.
[38] J. Pearl, "Reasoning with belief functions: An analysis of compatibility," *International Journal of Approximate Reasoning,* vol. 4, no. 5-6, pp. 363-389, 1990.
[39] L. A. Zadeh, "A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination," *AI magazine,* vol. 7, no. 2, pp. 85-85, 1986.
[40] R. R. Yager, "On the Dempster-Shafer framework and new combination rules," *Information sciences,* vol. 41, no. 2, pp. 93-137, 1987.
[41] L. Xiao, S. Boyd, and S. Lall, "A Scheme for Robust Distributed Sensor Fusion Based on Average Consensus," in *Information Processing in Sensor Networks, 2005. IPSN 2005. Fourth International Symposium on*, 2005.
[42] R. Olfati-Saber, "Distributed Kalman Filter with Embedded Consensus Filters," in *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference 2005*, Seville, Spain, 2005.
[43] G. Welch and G. Bishop, "An introduction to the Kalman filter," 1995.
[44] W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method," *Journal of Power Sources,* vol. 196, no. 23, pp. 10314-10321, 2011.
[45] P. Xu, Y. Deng, X. Su, and S. Mahadevan, "A new method to determine basic probability assignment from training data," *Knowledge-Based Systems,* vol. 46, pp. 69-80, 2013.
[46] O. Basir and X. Yuan, "Engine fault diagnosis based on multi-sensor information fusion using Dempster--Shafer evidence theory," *Information Fusion,* vol. 8, no. 4, pp. 379-386, 2007.
[47] J. W. Nilsson and S. A. Riedel, *Electric Circuits*. Prentice Hall, 2011.
[48] M. M. Begovic and A. G. Phadke, "Dynamic simulation of voltage collapse," in *Conference Papers Power Industry Computer Application Conference*, 1989: IEEE, pp. 336-341.
[49] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *Journal of mathematical sociology,* vol. 2, no. 1, pp. 113-120, 1972.
[50] Z. Wang, A. Scaglione, and R. J. Thomas, "Electrical centrality measures for electric power grid vulnerability analysis," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, 2010: IEEE, pp. 5792-5797.

APPENDIX

A Numerical Calculation Examples

A.1 Unweighted Eigenvector Centrality

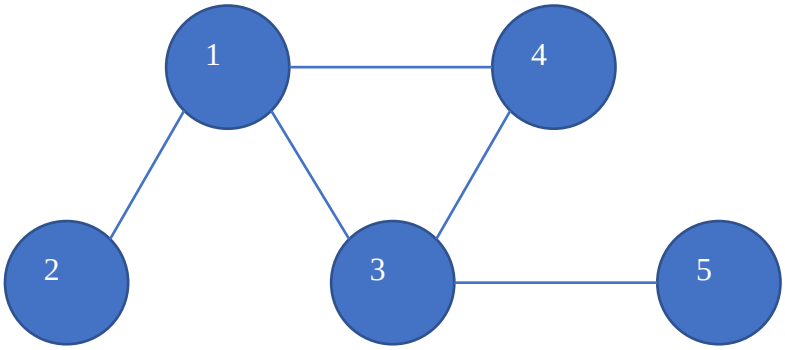


Figure 34: System topology for EC calculation example

The modified adjacency matrix is for nodes,

$$\hat{A} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

The beginning eigenvector is

$$e = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

The first iteration produces

$$\hat{A}e = \begin{bmatrix} 3 \\ 1 \\ 3 \\ 2 \\ 1 \end{bmatrix}$$

Normalized to the eigenvalue which is the maximum term of the eigenvector, $\lambda \equiv \max e$,

$$\lambda e = 3 \begin{bmatrix} 1 \\ 1/3 \\ 1 \\ 2/3 \\ 1/3 \end{bmatrix}$$

The error, ε, is the maximum difference between the sides of the eigenvector equation, or equivalently, the difference in the RHS or LHS calculation between two iterations:

$$\varepsilon = \max \left| \lambda e^{(i)} - \left(\hat{A}e\right)^{(i)} \right| = \max \left| \lambda e_k^{(i)} - \lambda e_k^{(i+1)} \right|$$

Since the maximum of e is always 1 therefore $\max(\lambda e) = \lambda$ and error term is simply the difference between two consecutive eigenvalues.

$$\varepsilon = \left| \lambda^{(i)} - \lambda^{(i+1)} \right|$$

Continued iterations are shown in Table 8: Unweighted EC steps for 5-node example Note that superscript (i) denotes the iteration of the variable and is not an exponent.

Table 8: Unweighted EC steps for 5-node example

Iteration, i	Operation	Node					λ	ε
		1	2	3	4	5		
0	$e^{(0)}$	1.000	1.000	1.000	1.000	1.000	1.000	
	$e^{(0)}\hat{A}$	3.000	1.000	3.000	2.000	1.000	3.000	2.000
1	$e^{(i)} = e^{(i-1)}\hat{A}/\lambda^{(i-1)}$	1.000	0.333	1.000	0.667	0.333		
	$e^{(i)}\hat{A}$	2.000	1.000	2.000	2.000	1.000	2.000	1.000
2	$e^{(i)} = e^{(i-1)}\hat{A}/\lambda^{(i-1)}$	1.000	0.500	1.000	1.000	0.500		
	$e^{(i)}\hat{A}$	2.500	1.000	2.500	2.000	1.000	2.500	0.500
3	$e^{(i)} = e^{(i-1)}\hat{A}/\lambda^{(i-1)}$	1.000	0.400	1.000	0.800	0.400		
	$e^{(i)}\hat{A}$	2.200	1.000	2.200	2.000	1.000	2.200	0.300
4	$e^{(i)} = e^{(i-1)}\hat{A}/\lambda^{(i-1)}$	1.000	0.455	1.000	0.909	0.455		
	$e^{(i)}\hat{A}$	2.364	1.000	2.364	2.000	1.000	2.364	0.164
5	$e^{(i)} = e^{(i-1)}\hat{A}/\lambda^{(i-1)}$	1.000	0.423	1.000	0.846	0.423		
	$e^{(i)}\hat{A}$	2.269	1.000	2.269	2.000	1.000	2.269	0.094

6	$e^{(i)} = e^{(i-1)}\hat{A}/\lambda^{(i-1)}$	1.000	0.441	1.000	0.881	0.441		
	$e^{(i)}\hat{A}$	2.322	1.000	2.322	2.000	1.000	2.322	0.053
7	$e^{(i)} = e^{(i-1)}\hat{A}/\lambda^{(i-1)}$	1.000	0.431	1.000	0.861	0.431		
	$e^{(i)}\hat{A}$	2.292	1.000	2.292	2.000	1.000	2.292	0.030
8	$e^{(i)} = e^{(i-1)}\hat{A}/\lambda^{(i-1)}$	1.000	0.436	1.000	0.873	0.436		
	$e^{(i)}\hat{A}$	2.309	1.000	2.309	2.000	1.000	2.309	0.017
9	$e^{(i)} = e^{(i-1)}\hat{A}/\lambda^{(i-1)}$	1.000	0.433	1.000	0.866	0.433		
	$e^{(i)}\hat{A}$	2.299	1.000	2.299	2.000	1.000	2.299	0.010

After 10 iterations the maximum ε between the steps is less than 0.01 and the calculation has reached the desired tolerance. The resulting centrality eigenvector is

$$C_e \approx e^{(9)} = \begin{bmatrix} 1 \\ 0.433 \\ 1 \\ 0.866 \\ 0.433 \end{bmatrix}$$

And the final eigenvalue is

$$\lambda = \max\left(\hat{A}e\right) \approx 2.309$$

Which satisfies the relationship

$$\lambda e = \hat{A} e$$

$$(2.309)\begin{bmatrix} 1 \\ 0.433 \\ 1 \\ 0.866 \\ 0.433 \end{bmatrix} \approx \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0.433 \\ 1 \\ 0.866 \\ 0.433 \end{bmatrix}$$

$$\begin{bmatrix} 1 \times 2.309 \\ 0.433 \times 2.309 \\ 1 \times 2.309 \\ 0.866 \times 2.309 \\ 0.433 \times 2.309 \end{bmatrix} \approx \begin{bmatrix} 0.433 + 1 + 0.866 \\ 1 \\ 1 + 0.866 + 0.433 \\ 1 + 1 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 2.309 \\ 1.000 \\ 2.309 \\ 2.000 \\ 0.433 \end{bmatrix} \approx \begin{bmatrix} 2.299 \\ 1.000 \\ 2.299 \\ 2.000 \\ 1.000 \end{bmatrix}$$

A.2 Dempster-Shafer Combination

In this example, the fusion is centered at sensor 1 which ascribes probability mass assignments m_1. The meaning of "centered" is that sensor only has contact with its neighbors at sensors 2, 3, 4. In this case, the focal frame is

$$\theta = \{h_1, h_2, h_3, h_4, h_5\}$$

The variable h_i represents the hypothesis that sensor i is the best measurement of the actual value of the item being measured. In the application of this framework to this study, the measurement would be the voltage collapse margin, M_p. The portion of the network under analysis could be represented by the following diagram (Figure 35),

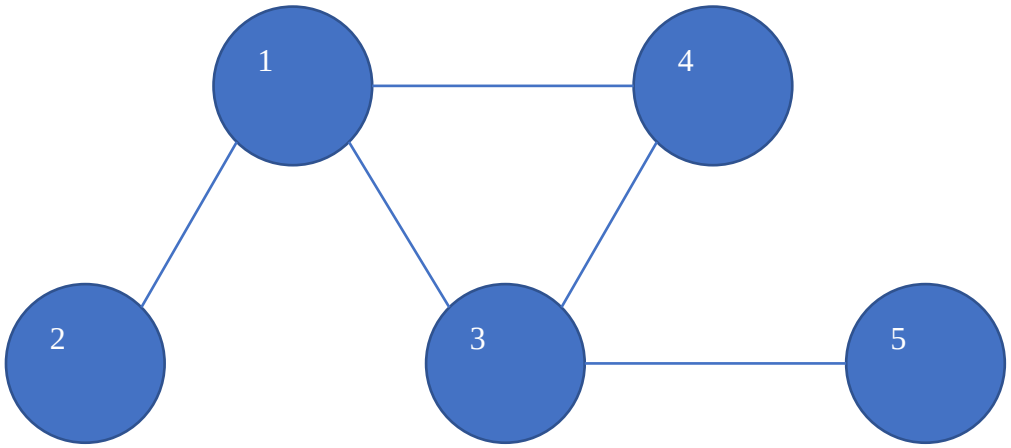


Figure 35:Representation of network for the sensor fusion example

Even though the fusion is centered at sensor 1, the focal frame θ contains a probability mass assignment for sensor 5 since it gains information about sensor 5 from sensor 3. The probability mass assigned to the focal frame at-large is the probability mass which has not been assigned to any specific hypothesis:

$$m(\theta) = m(h_1 \cap h_2 \cap h_3 \cap h_4 \cap h_5)$$

Note that the probability mass, $m(a)$ is the belief that one commits exactly to the subset $\{a|a \subset \theta\}$ and does not include any probability mass assignments $m(b)$ of any subsets b of a, $\{b|b \subset a\}$. Thus, the probability mass assignment to the focal frame θ is the probability mass not assigned to any subsets of θ, or, equivalently, the unassigned probability mass. It is not the total belief that is committed to θ; this would be expressed as $Bel(\theta) = 1$ which corresponds to

the familiar Bayesian result $P(\theta) = 1$ or the truism that we can state with 100% certainty that probability that at least one measurement in the set is the best measurement of the set.

The initial probability assignments are given in Table 9 below.

Table 9: Initial probability assignments

	m_1	m_2	m_3
h_1	0.1	0.4	0.1
h_2	0.1	0.5	0
h_3	0.1	0	0.2
h_4	0	0	0.4
θ	0.7	0.1	0.4

The form for an event, c, specified by the combination of evidence from A and B is

$$m_{12}(c) = K \sum_{h_i \cap h_j = c} m_1(h_i) m_2(h_j) .$$

The normalization factor, K, is the inverse of one minus the sum of the masses of the empty sets which result from the combination of all observations in the power set:

$$K^{-1} = 1 - \sum_{a_i \cap b_j = \emptyset} m_A(a_i) m_B(b_j)$$

If the evidence set is specified by a column vectors M_A and M_B, then the combination terms can be expressed as a matrix M_{AB} which is the vector product of $M_A\, M_B^T$. This is illustrated graphically in Table 10 below.

$$M_A\, M_B^T = M_{AB}$$

Table 10: Combination of m1 and m2

	$\boldsymbol{m_2(h_1) = 0.4}$	$\boldsymbol{m_2(h_2) = 0.5}$	$\boldsymbol{m_2(h_3) = 0}$	$\boldsymbol{m_2(h_4) = 0}$	$\boldsymbol{m_2(\theta) = 0.1}$
$m_1(h_1) = 0.1$	$m(h_1) = 0.04$	$m(\emptyset) = 0.05$	$m(\emptyset) = 0$	$m(\emptyset) = 0$	$m(h_1) = 0.01$
$m_1(h_2) = 0$	$m(\emptyset) = 0$	$m(h_2) = 0$	$m(\emptyset) = 0$	$m(\emptyset) = 0$	$m(h_2) = 0.01$
$m_1(h_3) = 0.1$	$m(\emptyset) = 0.04$	$m(\emptyset) = 0.05$	$m(h_3) = 0$	$m(\emptyset) = 0$	$m(h_3) = 0.01$
$m_1(h_4) = 0$	$m(\emptyset) = 0$	$m(\emptyset) = 0$	$m(\emptyset) = 0$	$m(h_4) = 0$	$m(h_4) = 0$
$m_1(\theta) = 0.7$	$m(h_1) = 0.28$	$m(h_2) = 0.35$	$m(h_3) = 0$	$m(h_4) = 0$	$m(\theta) = 0.07$

Since our hypotheses are non-overlapping and the uncertainty mass is the last element in the vectors, the normalization factor is the inverse of one minus the sum of all of the off-diagonal elements of the $N \times N$ matrix, M_{12}, which are not elements of the last row or column.

$$K^{-1} = 1 - \sum_{\substack{i=1;\ j=1 \\ i \neq ji}}^{N-1} m_1(h_i) m_2(h_j)$$

$$= 1 - (0.05 + 0.04 + 0.04 + 0.05) = 0.82$$

The probabilities assigned to each element are the corresponding diagonal plus the corresponding element of the last row and column:

$$m_{12}(h_i) = K^{-1}(M_{Ni} + M_{ii} + M_{iN})$$

$$m_{12(h_1)} = \frac{0.04 + 0.28 + 0.01}{0.82} \approx 0.402$$

$$m_{12}(h_2) = \frac{0.35 + 0.05 + 0.01}{0.82} \approx 0.5$$

$$m_{12}(h_3) = \frac{0.01}{0.82} \approx 0.0122$$

$$m_{12}(h_4) = 0$$

$$m_{12}(\theta) = \frac{0.07}{0.82} \approx 0.0854$$

Since the fusion process is associative, the output of the next sensor m_3 is fused with the output of the first combination m_{12}.

	$m_3(h_1) = 0.1$	$m_3(h_2) = 0$	$m_3(h_3) = 0.2$	$m_3(h_4) = 0.4$	$m_3(\theta) = 0.3$
$m_{12}(h_1) = 0.402$	$m(h_1) = 0.0402$	$m(\emptyset) = 0$	$m(\emptyset) = 0.0805$	$m(\emptyset) = 0.161$	$m(h_1) = 0.121$
$m_{12}(h_2) = 0.5$	$m(\emptyset) = 0.05$	$m(h_2) = 0$	$m(\emptyset) = 0$	$m(\emptyset) = 0.2$	$m(h_2) = 0.15$
$m_{12}(h_3) = 0.0122$	$m(\emptyset) = 0.0012$	$m(\emptyset) = 0$	$m(h_3) = 0.0024$	$m(\emptyset) = 0.0049$	$m(h_3) = 0.0037$
$m_{12}(h_4) = 0$	$m(\emptyset) = 0$	$m(\emptyset) = 0$	$m(\emptyset) = 0$	$m(h_4) = 0$	$m(h_4) = 0$
$m_{12}(\theta) = 0.0854$	$m(h_1) = 0.0085$	$m(h_2) = 0$	$m(h_3) = 0.0171$	$m(h_4) = 0.0342$	$m(\theta) = 0.0256$

The final output of the sensor fusion in this example is

$$K^{-1} \approx 0.402$$

$$m_{123}(h_1) \approx 0.421$$

$$m_{123}(h_2) \approx 0.373$$

$$m_{123}(h_3) \approx 0.0576$$

$$m_{123}(h_4) \approx 0.0849$$

$$m_{123}(\theta) \approx 0.0636$$

The results of this three sensor combination example are summarized in the table [TABLE]

	m_1	m_2	m_3	m_{123}
h_1	0.1	0.4	0.1	0.421
h_2	0.1	0.5	0	0.373
h_3	0.1	0	0.2	0.0576
h_4	0	0	0.4	0.0849
θ	0.7	0.1	0.4	0.0636