Cross-Domain Data Fusion On Distribution Network Voltage Estimation with D-S Evidence Theory

Yuanbin Zhu
Department of Electronic & Electrical Engineering
The University of Bath
Bath, the United Kingdom
ybz22@bath.ac.uk

Chenghong Gu
Department of Electronic & Electrical Engineering
The University of Bath
Bath, the United Kingdom
C.Gu@bath.ac.uk

Furong Li
Department of Electronic & Electrical Engineering
The University of Bath
Bath, the United Kingdom
F.Li@bath.ac.uk

***Abstract*— The Cyber-Physical system (CPS) is an emerging concept for realizing the system wholeness and the interplay of different network components in the electricity system with various embedded devices. However, the increasing penetration of embedded devices also brings severe data explosion and uncertainties to the voltage estimation process in practical power grid operations. Data fusion method has significant performance on improving the accuracy of state estimation on the deficient cross-domain dataset. This paper applies the data fusion method and D-S evidence theory to aggerate the information from various monitored devices in the distribution network and resolve the voltage estimation problem of the distribution network. Apart from conventional data fusion model, a two-stage D-S evidence data fusion framework is also proposed to improve the estimation accuracy and also quantify correlation factors between recorded parameters of monitored network devices and the overall network status of the whole distribution systems. This paper illustrates the feasibility and reliability of the proposed data fusion frameworks by a case study on an actual MV distribution system with deficient datasets and operation information of the main 33/11 kV transformer.**

Keywords—Cyber-physical systems, Data fusion, Decision making, Dempster-Shafer (D-S) evidence theory

I. Introduction

Energy decarbonizing has become a significant challenge for the power industry governments and researchers. For reducing the greenhouse gas emission, renewable power generation and the plug-in electric vehicles (PEVs) are promoted as the future direction for the power and transportation section. Meanwhile, they also put forward higher requirement at power supply quality [1].

Benefits from the development of CPS, system operators start to monitor and maintain the status of power system by various embedded monitoring devices installed at the different component in the power system [2]. The continuously increasing penetration of intelligent monitoring devices in the power system provides a huge amount of detailed information about the power system and allows system operators to have a comprehensive estimation on the overall condition of the whole power system.

However, the explosion of monitoring data also significantly increases the complexity of network analysis and the cost of the communication network. Also, as the restriction of installation cost and customer privacy, the distribution of embedded monitoring devices are normally concentrated at substations and centralize generation units, while other network components, such as distributed generators (DGs) and the load demand of user terminals become the blind point in the network [3], and thus brings great challenge on the state estimation of the whole network. Besides, at the scenario of lacking sufficient embedded monitoring devices, such as devices failure, un-upgraded old distribution, the collected information is normally located at different domains (e.g. Individual voltage, current and load database at different buses), and thus brings severe information gaps into the network analysis process.

In conventional network study, the state estimation study is a well-established tool for analyzing the status of electricity network [4]. In [5], a three-stage distributed state estimation model is proposed for the future distribution network under mixed measurement environment. However, the initialization of the proposed three-stage framework requires the complete data of network topology, measurement and parameter of the distribution system. In [6], a weighted least square-based distribution system state estimation approach is proposed to enhance the load pseudo measurement accuracy of the distribution system. But, the performance of dynamic state estimation framework is sensitivity to the completeness of input databases, and thus the cross-domain information has a significant impact on increasing the analysis complexity and uncertainty. [7] evaluate the suitability of various estimation methodologies for the estimating the overall state of distribution network, and indicate the performance of evaluated techniques rely on the detailed information with the known parameters such as noise factor.

Data fusion is an emerging analysis technique for combing data from multiple sources [8]. When analysisng the power system, conventional model-based methodologies normally require the complete database to establish the complete system model. Differently, data fusion method aggregats information from different sensors and extract multi-dimensional information for global decision-making. In [9], data fusion technique is adopted to aggregate various forecasts from different patterns of price signals. Similarly, data fusion has significant potential on analysing medium voltage (MV) and low voltage (LV) distribution networks with various uncertainties and data imperfection problem in grid integrated operation.

D-S evidence theory is a general analysis framework which allows reasoning information uncertainty. In [10], evidence theory is utilized to fuse probabilistic, possibilistic, and interval uncertainties in power flow analysis. Besides, D-

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

S evidence framework also has a significant advantage in evaluating the weights of different data inputs. Paper [11] proposes a multi-objective group decision-making condition-based maintenance model based on D-S evidence theory to consider the comprehensive weights of various estimations from different transmission and transformation equipment.

For solving the challenge of the distribution system state estimation on deficient cross-domain datasets, this paper proposes a cross-domain data fusion framework based on D-S evidence theory to estimate the overall voltage level of distribution networks with the deficient cross-domain database. Thereafter, it adopts the modified methodology [12] to amend the evidence source rather than modify the combination law for reducing the impact of conflict information in the D-S evidence framework [13]. Then, a two-stage fusion framework is designed for evaluating the weights of various parameters to fully exploit the advantage of evidence framework. Finally, the two-stage data fusion framework based on D-S evidence theory is illustrated on an actual 11kV MV distribution network with the deficient database from eight different transformers in the network.

The rest of this paper is organized into four sections. Section II introduces the D-S evidence theory and the conflict management method for improving fusion accuracy. Section III formulates the testing system, the conventional D-S evidence framework and the proposed two-stage D-S data fusion system. Section IV presents the case study of data fusion method on an actual MV distribution network with deficient transformer datasets. Section V concludes the feasibility and reliability of data fusion methodology and related further development.

II. Methodology

For the solving problem of the distribution system state estimate on the deficient cross-domain dataset, section II firstly illustrate D-S evidence methodology and related conflict management theory for fusing the scattered information from the cross-domain dataset

A. D-S Evidence Framework

Evidence theory is a well-established framework for sensor fusion. As a generalization of the Bayesian theory of subjective probability [14], its specific index, degree of belief, has a significant advantage in representing the correlation between observations and target information. In the D-S evidence framework, the degree of belief for input variables is demonstrated by the basic belief assignment (BBA) or basic probabilistic assignment (BPA). The general definition of the framework is:

$$m: 2^{\theta} \rightarrow [0,1], m(\emptyset) = 0, \sum_{A \subseteq \theta} m(A) = 1 \tag{1}$$

where θ is the identification framework, 2^{θ} is the power set of θ, m is the basic probability assignment function within θ, $m(A)$ is the BPA for evidence set A. For each evidence set, a plausibility function and a belief function are also proposed for assigning the upper and lower boundary of the BPA,

$$bel(a) \leq P(A) \leq pl(A) \tag{2}$$

$$bel(A) = \sum_{B|B \subseteq A} m(B) \tag{3}$$

$$pl(A) = \sum_{B|B \cap A \neq \emptyset} m(B) \tag{4}$$

Another significant advantage of D-S evidence framework is that it allows aggregating various independent cross-domain evidence in a general identification framework into a belief improved estimation for decision making. The Dempster combination law is formulated as (5),

$$m_{1,2}(A) = \begin{cases} \dfrac{\sum_{A_i \cap B_j \subseteq \theta} m_1(A_i) m_2(B_j)}{1-k}, & A \neq \emptyset \\ 0 \qquad , & A = \emptyset \end{cases} \tag{5}$$

$$k = \sum_{A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j) \tag{6}$$

$$i = 1, 2, \ldots, n_A$$
$$j = 1, 2, \ldots, n_B$$

where A and B are two independent evidence sets in the identification framework θ, A_i and B_i are the focal elements in evidence sets A and B respectively, k is the conflict coefficient between evidence sets A and B, n_A and n_B are the numbers of focal elements in evidence set A and B.

Referred to the D-S evidence framework, each type of parameters in each database from the various sensors on different network components (e.g. transformers, load or branches) can be treated as various focal elements in different databases. Meanwhile, the correlation between the target variable and various recorded parameters, for example, the voltage levels of unmonitored buses and monitoring data from surrounding network components, also can be treated as the relative degree of belief with evidence theory. Thereafter, the individual degree of belief from each monitoring parameters can be then normalized and aggregated with the Dempster combination rule, which can then be used to assess the unmonitored component in the power system.

B. Conflict Management

One significant feature of the D-S evidence framework is the global fusion for multiple evidence set, and thus, this characteristic also leads to that D-S evidence theory is very sensitive to the amount of conflict between different evidence set. To decrease the impact, the minority evidence sets with high conflict coefficient are modified based on [12],

$$m(A_i)'_{m(A) \cap m(B) = \emptyset} = \begin{cases} 10^{m(A_i) - \frac{1}{n_{A_i}}}, & m(A) < \dfrac{1}{n_A} \\ 10^{m(A_i) + \frac{1}{n_{A_i}}}, & m(A) \geq \dfrac{1}{n_A} \end{cases} \tag{7}$$

$$m(A_i) = \frac{m(A_i)'}{\sum_{i=1}^{n_i} m(A_i)} \tag{8}$$

III. Problem Formulation

In modern power system, the complexity of distribution network dataset, including load, DGs and branch connections, has been considered as a significant challenge on applying proper state estimation analysis for the network dispatch process. For the traditional state estimation method, the high requirement of the database completeness brings a significant restriction for its application on the distribution voltage estimation problem. With the D-S evidence theory, data fusion model can reduce the negative impact of information redundancy and deficiency in distribution network databases. This section illustrates and introduces various monitoring parameters from the modern power system database into the

D-S evidence-based data fusion model. Also, the two-stage D-S evidence data fusion framework is proposed based on the historical data of three-phase voltage and three-phase currents to provide decision making support for modern network operators.

A. Testing System and Input Parameters

Firstly, this study targets to evaluate the overall voltage status of the MV distribution network by estimating the voltage level of the far-end LV distribution network. Thus, this paper proposes an actual MV distribution system with vague branch interconnection for the case study of data fusion framework and

The database of the MV distribution network consists of eight 11kV/415V transformers, located at different buses. Besides, in the daily dispatch process, the system operators generally raise or reduce the tap ratio of the main transformer as the voltage of terminal LV distribution network reach the lower or upper bound of the voltage standard. Thus, this study selects the main 33kv/11kv transformer in the testing network as the contrast component for emulating the state of the global voltage level of the whole distribution system and verify the accuracy of the fusion results.

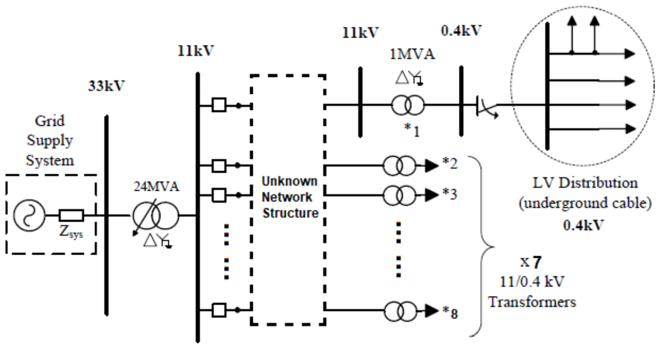

Figure 1. MV distribution network for case study

Each 11/0.4 kV transformer is connected to a 0.4kV LV distribution network, but the network structure of the interconnection branches among these eight buses are unknown. This paper selects the main 33/11 kV transformer as the target assessment component and targets to estimate the voltage level at the far end of the distribution network (i.e. 0.4kV LV distribution network). The three-phase voltages and three-phase currents of these eight transformers are selected as input variables of the data fusion framework.

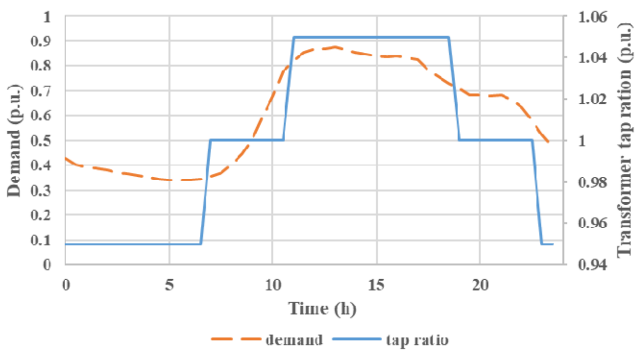

Figure 2. the variation of tap ratio and demand at a typical day

As shown in Table 1, the datasets of different parameters of theses eight transformers has different size, time length and deficiency level. For eliminating the data deficiency impact, the D-S evidence framework is introduce in this case rather than conventional state estimation method. Besides, the tap ratio of main transformer serves as the contrast component for the estimated voltage level from the data fusion framework.

TABLE I. THE NUMBER OF RECORDED PARAMETERS IN 3 PHASE SETS

Historical dataset	**The Number of Recorded Parameters** *(in 3 phase set)*		
	Voltage data (group)	***Current data (group)***	***Damaged data (group)***
Transformer 016	10614	10224	393
Transformer 028	6432	6384	48
Transformer 029	9936	10176	240
Transformer 030	1825	1728	97
Transformer 033	9985	9860	125
Transformer 034	6624	6720	96
Transformer 035	10620	10365	255
Transformer 079	9244	9165	79
Main Transformer	Contrast Component		

B. Traditional D-S evidence Framework

In the traditional D-S evidence model, the input parameters commonly consist of the same type of focal elements form similar evidence sets (i.e sensors) with the same correlation related to the assessment target. Thus, this method can directly aggerate the input information to provide a brief estimation of the target component.

To resolve the voltage estimation problem in the distribution network, this paper selects the voltage records from different transformers in the testing network as the input parameters of the conventional D-S evidence framework to estimate the voltage level of the far-end LV distribution system in the MV distribution network.

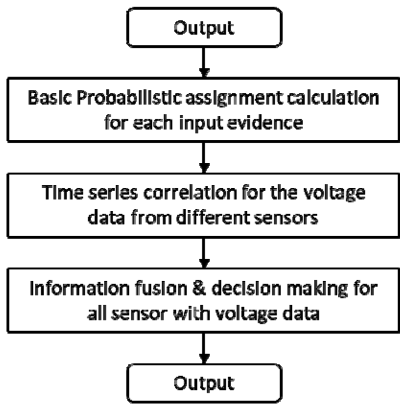

Figure 3. The flowchart of Traditional D-S evidence framework

As shown in Figure 3, the traditional data fusion framework is generally separated into two steps after collecting the input information from different sources. First, this system synchronizes the voltage data from different datasets based on their time-series information. Second, the system aggregates the new synchronized voltage dataset and provides a reference value for evaluating the overall voltage status of the whole MV distribution network.

C. Two-stage Evidence Fusion Framework

Traditional D-S evidence framework normally adopts equal weight for all input focal elements from the same evidence set. For power system analysis, the spatial distribution of different buses and the temporal relation of the recorded information from different parameter have a significant impact on the data correlation. Thus, this paper adopts a two-stage fusion framework to quantify the correlation between different input evidence and target estimation variables. Generally, the overall data fusion framework is divided into two stages: the calculation of correlation factors and overall estimation for the target variable.

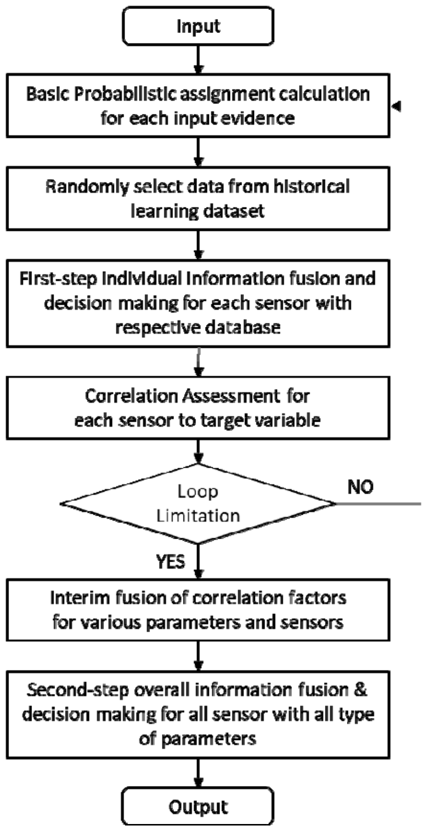

Figure 4. The flowchart of two-stage D-S evidence framework

The first-stage fusion targets to calculate the correlation between each sensor (i.e. evidence set) to the assessment target. It firstly applies individual data fusion on each transformer dataset with D-S evidence theory to generate a brief estimation to the target voltage level. Then, the first-stage data fusion calculates the correlation coefficients based on the individual data fusion result for each transformer. The correlation calculation is based on Pearson's correlation coefficient equation,

$$\rho_{XY} = \frac{cov(X,Y)}{\sigma X \sigma Y} \tag{9}$$

$$m_{12}(A) = \frac{\sum_{A_i \cap B_j \subseteq \theta} \rho_A m_1(A_i) \rho_B m_2(B_j)}{1-k} \tag{10}$$

where ρ_A and ρ_B are the correlation coefficient between evidence set *A* and the target variable. After collecting enough data for correlation factor quantification, an interim fusion is then proposed to calculate improved correlation factors for each sensor to quantify the impact of different evidence set in the final decision-making process.

Thereafter, the second-stage data fusion is then proposed for estimating the real-time condition of the target variable with the real-time monitoring data. In this stage, all data from different sensors will be globally aggregated with the improved correlation factors from the first-stage data fusion process. Finally, the decision-making process will be executed based on the combined degree of belief (i.e. data fusion result) from the second-stage data fusion process.

Overall, the first-stage data fusion is similar to the conventional data fusion framework. It firstly generates the fusion estimation for each transformer and then calculates the correlation coefficient between each 11/0.4 kV transformer to the target LV distribution network. The second-stage data fusion is an adjusted D-S evidence-based data fusion framework with the correlation coefficient. The introduction of correlation coefficient targets to reduce the negative impact of the deficient cross-domain information and also take the full advantage of the data size of the cross-domain database.

IV. CASE STUDY

For two-stage data fusion model, one primary object of the two-stage data fusion framework is to exploit the correlation between cross-domain information which cannot be quantified in the traditional state estimation method. Thereafter, this paper resolves the voltage estimation problem with deficient cross-domain datasets. For conventional D-S evidence framework, it serves as the comparison for the proposed two-stage model. This section illustrates the case study in two parts, the conventional D-S framework and two-staged data fusion framework (including the first-stage data fusion and the second stage data fusion).

A. Conventional D-S evidence

As stated in section III, the Conventional D-S evidence framework estimates the voltage level of the far-end LV distribution network through fusing the voltage data from various separated transformers in the distribution power system.

In this case study, the voltage level of then far-end LV distribution network is generally divided into three domains based on UK voltage standard: High (1.02-1.06 p.u.), Medium (0.98-1.02 p.u.), Low (0.94-0.98 p.u.). Besides, this study also chooses -1, 0 and 1 as the output symbol for Low, Medium and High voltage level of the LV distribution network respectively. Figure 5 and Table II shows the fusion result of the conventional D-S evidence framework.

TABLE II. FUSION RESULT OF TWO-STAGE D-S EVIDENCE FRAMEWORK

D-S Evidence Framework	*Total Fusion Timestep*	*Correct Fusion Result*	*Overall Consistency*
Data Fusion Result	144	121	84.02%

As shown in Table II and Figure 5, the variation of fusion result and the tap ratio of the main 33/11 kV main transformer have a high consistency of 84.02%. This phenomenon shows that the D-S evidence theory has good performance in the voltage estimating of the distribution network. Besides, the D-S evidence framework also shows its sensitivity and foreseeability on the sudden demand change, especially at around 7:00 a.m., which means the data fusion system detects the voltage drop or demand change before operation record in the recorded database. Generally speaking, the results show the feasibility and reliability of the D-S evidence theory on solving the voltage estimating problem of distribution network.

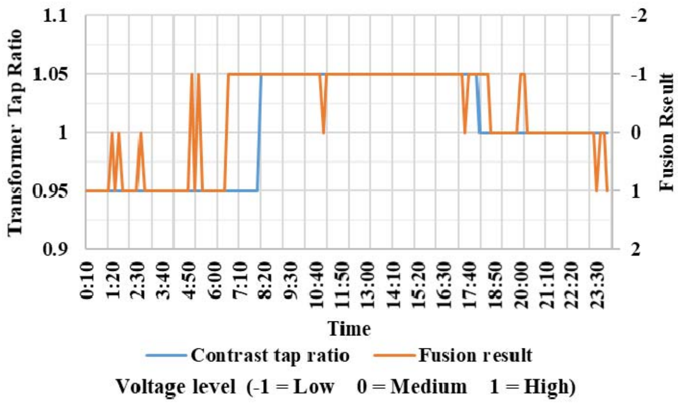

Figure 5. Combined degrees of belief for transformation ratio

B. *First-stage Data Fusion*

Differently, the proposed D-S evidence-based data fusion framework is formulated in two stages. The first-stage data fusion targets to evaluate the correlation between each monitored transformers and the target voltage level of the far-end LV distribution network based on the historical datasets. Table III below shows correlation coefficients between eight 11/0.4 kV transformers and target LV distribution network.

TABLE III. CORRELATION COEFFICIENT OF INDIVIDUAL TRANSFORMER

Transformer No.	**Correlation Coefficient**		
	Maximum	***Average***	***Minimum***
Transformer 016	0.948	0.892	0.819
Transformer 028	0.928	0.844	0.560
Transformer 029	0.950	0.894	0.850
Transformer 030	0.912	0.806	0.689
Transformer 033	0.947	0.887	0.543
Transformer 034	0.929	0.843	0.548
Transformer 035	0.914	0.864	0.671
Transformer 079	0.913	0.825	0.721

As shown in Table III, the maximum correlation coefficients of different transformers (or connected buses) to the target terminal LV distribution network are all above 0.9. The results of the correlation calculation show the high consistency between these eight LV buses to the target LV distribution network. Additionally, this phenomenon also shows the possibility that there is no large distributed generator connected between these LV buses and the slack bus because of the high correlation between the target voltage level and the voltage data from these eight distributed 11/0.4 kV transformers. Furthermore, the transformer 030 has the lowest maximum and average correlation coefficient to the main transformation ratio, and thus its connected 0.4 kV distribution network has the smallest impact to the target LV distribution network among all the eight LV distribution transformers.

C. *Second-stage Data Fusion*

After evulating the correlation relationship between various input parameters and target voltage level, the second-stage data fusion aggregates all the data from the eight transformers on a typical day to estimate the transformation ratio of the 33/11 kV main transformer in the MV distribution network.

The combined degree of belief shows the output tendency of the data fusion system on three voltage levels. As shown in figure 6, the combined degree of belief varies sharply at around 9:00 a.m. and 19:00. These rapid changes are cause by the sudden load increase of starting daily work and sudden demand decrease after peak hours. In other words, the combined degree of belief from D-S data fusion system have reliable performance on detecting the sudden voltage variation in the MV distribution network. Then, the second-stage data fusion will provide a auxiliary estimation of the voltage level to system operators for their decision making process.

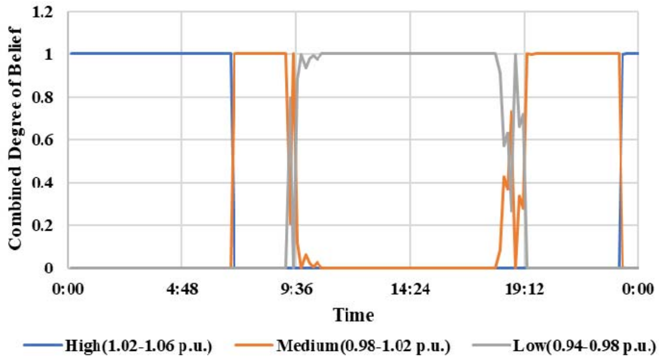

Figure 6. Combined degrees of belief for transformation ratio

As stated in section III, the operation information of the main transformer shows the overall voltage level of the whole distribution network to son extents. The system operators normally raise the tap ratio of the main transformer when the voltage level of terminal LV distribution is sliding out of the lowest allowance range, which also means the overall voltage level of the whole MV distribution network is reaching its lower bound.

Similarly, the network operators will also reduce the main transformation tap ratio when the voltage reaches the higher allowance value. Thus, the 33/11kV main transformer serves as the contrast component for the voltage estimation from the D-S data fusion system in this case study. Table IV and Figure 7 shows the fusion results of the transformation ratio of the 33/11 kV main transformer, with the curve transformation ratio variation in a typical 24-hour period of the MV distribution network.

TABLE IV. FUSION RESULT OF TWO-STAGE D-S EVIDENCE FRAMEWORK

D-S Evidence Framework	***Total Fusion Timestep***	***Correct Fusion Result***	***Overall Consistency***
Data Fusion Result	144	130	90.28%

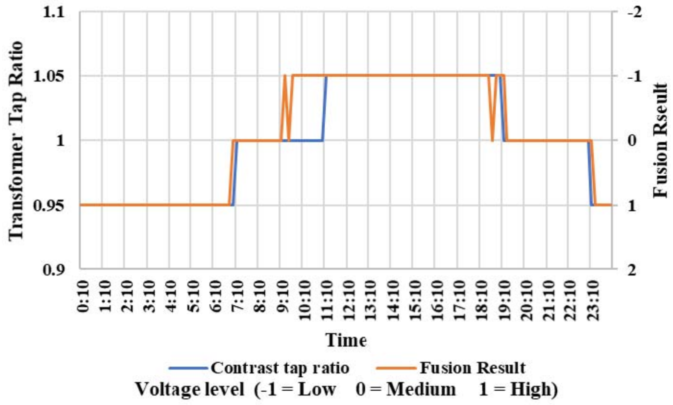

Figure 7. Fusion Result for 33/11 kV transformation ratio

As shown in Table. IV, the overall accuracy of the two-stage D-S evidence data fusion framework has reached 90.28% with 130 correct estimations in 144 timesteps. The consistency of the data fusion system improves 6.26% after introducing additional variables from transformer datasets.

More importantly, as shown in Figure 7, the variation curve of the fusion result remains stable except the sudden demand change at around 9:00 and 19:00, which means the reliability of the data fusion system has significant improvement as output remains stable when there is no sudden change in the MV distribution network.

Overall, the data fusion system of two-stage D-S evidence framework has a significant performance on evaluating the voltage level of the distribution system under a deficient database with various cross-domain measurements. Moreover, the two-step D-S evidence-based data fusion framework retains its foreseeability on sudden voltage changes from conventional D-S evidence framework. Thus, the two-stage D-S evidence framework can provide accurate and reliable information to the distribution system operators for their dispatch process. For example, the network should consider increasing the ratio of main 33/11 kV transformer to ensure the voltage level of the far-end LV distribution network when the system fusion result gradually slides from 1 to -1.

V. Conclusion

As analyzed, the reconfigured data fusion framework has a significant performance in evaluating the voltage of the distribution network and provide early warning to the sudden voltage and demand changes. Through introducing various input variable and Pearson's correlation coefficient, it can provide reliable correct fusion for monitoring the voltage variation of the MV distribution network. Based on the estimation, the network operator can evaluate the overall status of distribution networks and make early dispatch planning for the distribution devices. Meanwhile, the variation of the combined degree of belief also reflects the sharp load change in the distribution network, such as the peak demand period. Moreover, based on the correlation factor from second-stage data fusion, network operators can briefly assess the correlation between different network component (i.e. buses, branches, transformers) with unknown branches interconnection in the same MV distribution system.With more cross-domain input parameters, the performance of the proposed two-stage D-S evidence framework has significant potential on estimating the target parameter and provide reliable support to system operation for distribution network operations

References

1. Kelly, L., A. Rowe, and P. Wild. *Analyzing the impacts of plug-in electric vehicles on distribution networks in British Columbia*. in *2009 IEEE Electrical Power & Energy Conference (EPEC)*. 2009.
2. Karnouskos, S. *Cyber-Physical Systems in the SmartGrid*. in *2011 9th IEEE International Conference on Industrial Informatics*. 2011.
3. *Industrial Problems Booklets*. 2019, University of Cambridge.
4. Primadianto, A. and C. Lu, *A Review on Distribution System State Estimation*. IEEE Transactions on Power Systems, 2017. **32**(5): p. 3875-3883.
5. Kong, X., et al., *Three-Stage Distributed State Estimation for AC-DC Hybrid Distribution Network Under Mixed Measurement Environment*. IEEE Access, 2018. **6**: p. 39027-39036.
6. Liu, Y., J. Li, and L. Wu, *State Estimation of Three-Phase Four-Conductor Distribution Systems With Real-Time Data From Selective Smart Meters*. IEEE Transactions on Power Systems, 2019. **34**(4): p. 2632-2643.
7. Singh, R., B.C. Pal, and R.A. Jabr, *Choice of estimator for distribution system state estimation*. IET Generation, Transmission & Distribution, 2009. **3**(7): p. 666-678.
8. Castanedo, F., *A Review of Data Fusion Techniques*. TheScientificWorldJournal, 2013. **2013**: p. 704504.
9. Darudi, A., M. Bashari, and M.H. Javidi, *Electricity price forecasting using a new data fusion algorithm*. IET Generation, Transmission & Distribution, 2015. **9**(12): p. 1382-1390.
10. Luo, J., L. Shi, and Y. Ni, *Uncertain Power Flow Analysis Based on Evidence Theory and Affine Arithmetic*. IEEE Transactions on Power Systems, 2018. **33**(1): p. 1113-1115.
11. Wang, H., et al., *Research on Multiobjective Group Decision-Making in Condition-Based Maintenance for Transmission and Transformation Equipment Based on D-S Evidence Theory*. IEEE Transactions on Smart Grid, 2015. **6**(2): p. 1035-1045.
12. WANG Yang, L.Y., ZHANG Qian, *Conflict Solution in D-S Evidence Theory*. Journal of jiangnan university, 2015(14(03)): p. 269-272.
13. Lefevre, E., O. Colot, and P. Vannoorenberghe, *Belief function combination and conflict management*. Information Fusion, 2002. **3**(2): p. 149-162.
14. Shafer, G., *Dempster-Shafter theory*. 2002.
15. Fusco, F., S. Tirupathi, and R. Gormally. *Power systems data fusion based on belief propagation*. in *2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe)*. 2017.