WAMS/SCADA Data Fusion Method Study Based on Time-Series Data Correlation Mining

Lijin Zhao1, Liang Huang1\*, Qiansu Lv1, Tao Yang1, Daqian Wei2

1. Electric Power Research Institute of Guizhou Power Grid Co., Ltd, GuiZhou, 550000, China
2. School of Electrical Engineering, Wuhan University, Wuhan, 430072, China

lianghuang@gmail.com

**Abstract:** Hybrid measurement state estimation of WAMS data and the SCADA system is an effective method to improve the traditional state estimation. However, as the WAMS data and the SCADA data belong to different systems, there are great differences between them. To solve this problem, WAMS/SCADA data fusion method based on the correlation mining of time-series data is proposed in this paper. Firstly, WAMS/SCADA correlation estimation is done with the derivation of Pearson correlation coefficient., and then, solving the function model for the time difference issue and the alignment problem of correlation curves. After that, analyzing the measurement precision by considering the measurement weight and calculate the matrix of time series data weight to complete the optimization for the measurement precision. Finally, form the effective fusion scheme based on the correlation of timing data. Simulation results on the IEEE 118 nodes system, with set a comparison of different hybrid measurement state estimation and different state estimation algorithm, effectiveness and stability of the proposed method has been proved.

**Keywords:** Time-series data; Correlation mining; WAMS/SCADA data fusion.

1. Introduction

With the Power Grid becoming smarter and more integrative, real-time data transmission and analysis in Power Grid is much more important. SCADA plays an important role in traditional power system analysis for a long time [1]. As the WAMS proposed and perfected, it can provide a new method of monitoring and analyzing of Power Grid. Taking no account of time delay, the WAMS is able to monitor measurement data of the whole Power Grid and provide unprecedented data stream to keep the Power Grid safe and stable [2-5]. However, it is difficult to analyze measurement data, make decisions and become a single reliable system source within a short time because of the inadequacy of deployment of WAMS and high-speed data sources. Therefore, hybrid measurement data of SCADA and WAMS based on response is an important on-line method to estimate and analyze the Power Grid.

Because of the differences between WAMS and SCADA’s technology platform, they are different in component, precision, transmission time delay and refresh rate. At present, Researches mainly focus on building the correlation constraint between the WAMS and the SCADA and hybrid measurement state estimation by nonlinearity state estimation or OLS, or simply improving PMU measurement precision [6]. This method faces trade-offs with measurement data to reduce differences and enhance precision, which ignores the integrity of data largely and can make good use of the WAMS/SCADA data.

Therefore, by the method based on time-series data correlation mining to evaluate correlation of data, the WAMS/SCADA data fuses according to its correlation to make the best use of data after curve registration or curve alignment [7]. The method based on time-series data correlation mining has been applied in many research field of subjects. For example, it can study stream flows, temperature and precipitation when predicting flood disaster to improve precision. It can also propose advice on inflation and economic trend by analyzing CPI and GDP. What’s more, it is able to locate the earthquake and earthquake scale according to wave sequence in different places. The object of time-series data correlation analysis is heterogeneous data which is from different sources or different property. For example, the WAMS/SCADA data in this paper should be evaluated its correlation before regression analysis [8].

WAMS/SCADA data fusion method based on time-series data correlation mining is proposed in this paper. The method firstly determines the correlation coefficient and evaluates its own correlation. Then solv the function model by time-series data curve alignment. Finally completing blended data fusion.

1. WAMS/SCADA Correlation Evaluation
2. Correlation Coefficient Derivation

WAMS data and SCADA data are stable between two short periods. Before determining correlation coefficient, the upper and lower limit of correlation coefficient is needed. Pearson correlation coefficient proposed in this paper is widely used in weighing sequence correlation.

Assume that there are two correspondent groups of heterogeneous data:

 (1)

Accord with: 

Then correlation coefficient of sample is:

 (2)

 is the average of sample ,is the average of sample 

When , simply getting the probability density function of . After transferring, getting asymptotic distribution:

 (3)

From formula (4),  is monotonic increasing function, its upper and lower limit of correlation coefficient is:

When :

 (4)

When :

 (5)

1. Determination of WAMS/SCADA Correlation Coefficient

According to the derivation of upper and lower limit of correlation coefficient, WAMS/SCADA correlation coefficient can be showed with:

 (6)

In last formula,  is the correlation coefficient of WAMS measurement data  and SCADA measurement data .  is the initial moment of measurement data in some period.  is the cut-off moment of measurement data in some period.  is average of SCADA measurement sample.  is average of WAMS measurement sample.

Then gett the cross covariance function of WAAMS/SCADA measurement data:

 (7)

1. Method of Correlation Determination

To describe the correlation characteristic of sequence, firstly giving the definition of time-lag series. Assume that there are two groups of time-series data:

 (8)

Defining the following sequence is the time difference series:

 (9)

 (10)



Therefore, it is able to determine time-series data correlation of time difference sequence now.

 is a series of correlation coefficient vector, which is combined with the correlation coefficient of WAMS/SCADA measurement data according to sequential permutation. When WAMS/SCADA’s initial moments  is synchronous, it is the most correlative moment.

1. Method of Time-series Data Curve Alignment
	1. **Curve Alignment Model Based on Correlation Coefficient**

Correlation coefficient of the WAMS/SCADA time-series data helps to evaluate itself correlation. If the series is correlative and existing time deviation, differences in timer shaft should be removed by curve alignment. As for heterogeneous data, non-dimensional criterion is needed to align curve alignment combined with heterogeneous data.

Pearson correlation coefficient is a non-dimensional way to describe the correlation or similarity between series. However, continuity function needs to be described by inner product. Completing curve alignment of heterogeneous data by building the following function:



 (11)

 is the function after aligning.

* 1. **Solving Model**

References [9-11] use EM algorithm to solve the problem of curve alignment optimization. But when the dimension of parameters is much more, it is difficult to get a satisfied solution. To solving the problem, objective function in this paper is used as expectation of likelihood function in EM algorithm and Generalized Expectation Maximization (GEM) is used to solve the model [12-15].

Main steps are listed:

1. Input WAMS/SCADA data in the same period.
2. Initialize time difference vector  , and get the permitted errors of iteration.
3. Make time-series data become functions, function models can reference formula (13)
4. Get time difference vectors by General expectation maximization.
5. Redo step 3) and step 4) until convergence.
6. WAMS/SCADA Correlation Data Fusion

By derivation of correlation coefficient and curve alignment of WAMS/SCADA data, most of time-series data has been disposed, but the weight of time-series data depends on measurement precision [16]. The WAMS/SCADA hybrid measurement state estimation precision depends on device measurement precision and time synchronism of hybrid measurement data:

 (12)

 is global error of measurement data.  is the error caused by time synchronism.  is the error from system.

The error  caused by synchronism is:

 (13)

 is the gradient of measurement data. And  is the deviation between measurement moment and hybrid measurement data synchronism moment.

With knowing the system error, global error can be gotten by solving the deviation between measurement moment and hybrid measurement data synchronism moment to work out the precision of mixed measured state estimation [17-19]. WAMS data has better time scale characteristics. Thus,  is gotten by comparing WAMS time scale with the most correlative moment [20, 21]. Every measurement data has its own time delay , which has the following probability density:

 (14)

is variance of . is hybrid measurement data synchronism moment.

During precise calculation, errors of devices and time synchronism have no effect on each other [22-24]. So the variance of global error can be described with:

 (15)

1. Simulation Case
	1. **Simulation System and Estimation Indexes**

In this paper, the following indexes are used to evaluate:

 (16)

 (17)

Formula (19) shows correlation fusion of time-series data on moment k. Formula (20) shows correlation fusion of time-series data during the whole time.

 is true value of measurement  on moment  in sequence.  is estimated value of measurement  on moment  in sequence. N is dimension of measurement vectors. M is the number of measurement.

Case system use IEEE118 node bus system showed in Fig. 1. The SCADA measurement data used in case contains all active power and reactive power between lines in power system and all voltage amplitudes. The measurement data measures up Gaussian distribution. Its standard deviation is 0.02 and error mean is 0. PMU are equipped in Bus 2, 5, 8, 11 and 14. PMU contains bus voltage vectors and all outlet current vectors. The amplitude measurement measures up Gaussian distribution. Its standard deviation is 0.005 and error mean is 0. Phase angle measurement also measures up Gaussian distribution. Its standard deviation is 0.002 and error mean is 0.



**Fig. 1.** IEEE 118 nodes system

* 1. **Data Fusion Simulation Experiment**

There are three cases in the simulation. They all experience stabilization period, disturbance period and returning stabilization period.

Case 1: Only use SCADA data to estimate state.

Case 2: Estimate state without WAMS/SCADA data

Case 3: Estimate state after WAMS/SCADA data fusion

The final results are showed with estimated standard deviation after state estimation. Case 3 use the correlation coefficient in some period to make that  is the datum. When they are not synchronous, the differences of data’s stabilization will be removed by solving curve alignment function.

The simulation result is showed as Fig. 2:



**Fig. 2.** Data fusion simulation results

From the figure, firstly, simulation data’s precision is terrible and curve fluctuates when simply using SCADA data to estimate state. In Case 2, although state estimation can keep satisfied precision in stabilization period, the curve fluctuates when disturbing. In Case 3, state estimation keep satisfied precision both in stabilization period and disturbance period after correlation fusion. State estimation has been improved after time-series data correlation fusion.

**Table 1.** Different measurement measuring mixed state estimation results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Measurement time difference/s | 1 | 0.9 | 0.8 | 0.7 | 0.6 |
| CASE 1ρ/10-2 | 8.98 | 8.89 | 8.61 | 8.34 | 7.99 |
| CASE 2ρ/10-2 | 4.27 | 4.27 | 4.27 | 4.27 | 4.27 |
| Measurement time difference /s | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 |
| CASE 2ρ/10-2 | 7.73 | 7.52 | 7.14 | 6.73 | 6.25 |
| CASE 3ρ/10-2 | 4.26 | 4.27 | 4.27 | 4.26 | 4.27 |

From Table 1, estimation precision declines when time difference of the WAMS/SCADA measurement data increases. However, in Case 3, estimation precision almost keeps invariable and data keeps stable.

* 1. **Comparison Among Cases**

Two types of hybrid measurement estimation algorithm are listed to compare with the algorithm in the paper:

Case 1: Based on nonlinear state estimation.

Case 2: Based on nonlinear and linear state estimation.

Case 3: Based on time-series data correlation data fusion state estimation.

The final results are described with estimated standard deviation in three cases.

The results are showed as Fig. 3.

****

**Fig. 3.** More comparison algorithm simulation results

From the results, although the stability of using single nonlinear state estimation is good, precision is lower than the other two cases. Using mixed state estimation has improved a lot, but its stability is not as good as Case 3.

**Table 2.** Three kinds of simulation examples estimated standard deviation

|  |  |  |  |
| --- | --- | --- | --- |
| BUS | CASE 1 | CASE 2 | CASE 3 |
| Standard estimate difference /10-3 | Standard estimate difference /10-3 | Standard estimate difference /10-3 |
| 1 | 0.8020 | 0.5545 | 0.5512 |
| 2 | 1.0026 | 0.6975 | 0.6975 |
| 3 | 0.9163 | 0.6308 | 0.6324 |
| 4 | 0.9441 | 0.6200 | 0.6205 |
| 5 | 0.8511 | 0.5644 | 0.5645 |
| 6 | 0.8806 | 0.6300 | 0.6267 |
| 7 | 0.9507 | 0.6835 | 0.6801 |
| 8 | 1.0514 | 0.6665 | 0.6612 |
| 9 | 0.8672 | 0.6520 | 0.6555 |
| 10 | 0.9535 | 0.6490 | 0.6501 |
| 11 | 0.9144 | 0.5795 | 0.5880 |
| 12 | 0.8738 | 0.5961 | 0.5961 |
| 13 | 0.8441 | 0.5660 | 0.5643 |
| 14 | 0.8863 | 0.5205 | 0.5201 |

From table 2, it can be seen that data fusion based on time-series data correlation mining has better effective estimation than other traditional algorithm in stability or disturbance period.

1. Conclusion

Hybrid measurement data state estimation improves the state estimation which only uses SCADA data with WAMS data. However, there is not an effective scheme to solve the problem of state estimation based on WAMS/SACDA data. The scheme of time-series data correlation fusion is proposed in this paper:

1. Pearson correlation coefficient function is introduced in this paper. Time-series WAMS/SCADA data correlation estimation is done by determining upper and lower limit and derivation of correlation coefficient.
2. In the case of time difference of time-series curves, optimizing data is fused by building and solving model function.
3. Calculating the matrix of time-series data weight to complete an effective scheme based on time-series correlation by analyzing measurement data.
4. The results are verified by IEEE 118 nodes system. The stability has improved a lot than other schemes.

Acknowledgment

This project is supported by Key Projects of China Southern Power Gird (GZ2014-2-0049)

References

1. Phadke A G. Synchronized phasor measurements in power systems[J]. IEEE Computer Applications in Power, 2013, 6(2): 10-15.
2. Phadke A G, Pickett B, Adamiak M，et al. Synchronized sampling and phasor measurements for relay and control[J]. IEEE Trans. on Power Delivery, 2014, 9(1): 442-452.
3. Yu Qinjie, Wang Xiaoru, You Jiaxun, et al. Equality constraints two-step state estimation model based on phasor measurements[J]. Power System Technology, 2007, 31(10): 84-88 (in Chinese).
4. Qin Xiaohui, Bi Tianshu, Yang Qixun. A new method for hybrid nonlinear state estimation with PMU[J]. Automation of Power Systems, 2017, 31(4): 28-32 (in Chinese).
5. Liu Jinfeng, Wang Shuyang. Survey on applications of wide area measurement system in power system analysis[J]. High Voltage Engineering, 2015, 33(7): 182-185.
6. Duan Jundong, Sun Yankai, Yin Xiugang. Voltage stability’ online prediction using WAMS[J]. High Voltage Engineering, 2009, 35(7): 1748-1752.
7. Ding Junce, Cai Zexiang. Wang Keying. Mixed measurements state estimation based on WAMS[J]. Proceedings of the CSEE, 2016, 26(2): 58-63 (in Chinese).
8. Adelfio G, Chiodi M, D’Alessandro A, Luzio D, D’Anna G, Mangano G. Simultaneous seismic wave clustering and registration[J]. Computers & Geosciences, 2012, 44: 60-69.
9. Ye Limin, Keogh E. Time series shapelets: A new primitive for data mining[C]. Proc. of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2009. 947-956.
10. Liu Xin, Yang MCK. Simultaneous curve registration and clustering for functional data[J]. Computational Statistics & Data Analysis, 2009, 53(4): 1361-1376.
11. Yu Erkeng, Power system state estimation[M]. Beijing: China Water Power Press, 1995: 62-64 (in Chinese).
12. State Grid Corporation of China. Technical specification for WAMS[S]. Beijing: State Grid Corporation of China, 2006 (in Chinese).
13. Power system relaying committee of the IEEE power engineering society. IEEE standard for synchrophasors for power systems[S]. New York: The Institute of Electrical and Electronics Engineers, Inc, 2005.
14. Liu Xioing, Yang Ming. Simultaneous curve registration and clustering for functional data[J]. Computational Statistics & Data Analysis, 2019, 53(4): 1361-1376.
15. Li Dalu, Li Rui, Sun Yuanzhang. Data compatibility analysis of WAMS/SCADA hybrid measurements state estimation[J]. Proceedings of the CSEE, 2010, 30(16): 60-66.
16. You Jiaxun, Huang Bin, Guo Chuangxin, et, al. State estimation using SCADA and PMU hybrid measurements[J]. High Voltage Engineering, 2009, 35(7): 1765-1796.
17. Wu Xing, Liu Tianqi, Li Xingyuan, Li Congshan. Optimal configuration of PMU based on data compatibility of WAMS/SCADA and improved FCM clustering algorithm[J]. Power System Technology, 2014, (03): 756-761.
18. Liu Daowei, Song Dunwen, Wang Hongying. Voltage stability online evaluation system based on WAMS and EMS[J]. Power System Technology, 2014, (07): 1934-1938.
19. Zhou Hong, Li Qiang, Lin Tao. Power system disturbance and operation identification based on WAMS Electric Power[J]. Automation Equipment 2011, (02): 7-11.
20. Li Hong. Research on some issues of power system state estimation based on WAMS[D]. North China Electric Power University (Beijing), 2010.
21. Ji Luyu , Wu Junyong , Zhou Yanzhen. Transient stability prediction of power system based on WAMS characteristic of perturbed voltage trajectory clusters[J]. High Voltage Engineering, 2015, (03): 807-814.
22. Wang Huaibao, Shen Hong, Zhang Ying. Grid synchronization method based on improved adaptive filter[J]. High Voltage Engineering 2014, (11): 3636-3641.
23. Yang Jun, Liu Pei, Hu Wenping. Method for predicting failure rate of power transmission equipment based on cloud theory[J]. High Voltage Engineering, 2014, (08): 2321-2327.
24. Hu Jun, Cai Jingwen, He Jinliang. Factors impacting self-healing overvoltage of smart distribution network connected with micro grids[J]. High Voltage Engineering, 2014, (05): 1559-1566.