International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-1S3, December 2019

Methods of Missing Data Handlingin One Shot Response based Power System Control

Niraj Dahal, Steven M. Rovnyak

Abstract*: *The work done in this paper addresses various methods of handling missing phasor samples obtained from power flow simulations using DSA tools like TSAT and PSAT. Pseudorandom numbers in MATLAB are used to simulate 0-10% of missing samples and are recovered using different extrapolation techniques. After recovery, samples are subjected to decision trees to assess the performance of one shot stabilizing controls like in [1], [2].The power system model used is the 176 bus model of Western Electrical Coordinating Council (WECC).

Keywords: Decision trees, missing data handling, one shot control, transient stability

I. INTRODUCTION

A wide area monitoring system technology employs phasor measurement units (PMUs) across the network to acquire electrical phasors to a central location. These phasors collected in nearly real time are then utilized for monitoring, protection and control of the power system network as explained in [3]. Advent of real-time PMU data acquisition technology has provided remedies to different power system stability issues by continuously analyzing swings of the measured variables. However, there is a risk that the data collected from PMUs might undergo quality issues [4] like being non-numeric numbers, or noisy values or simply missed samples. Reference [5] mentions that on an average, 5-10% of missing samples have been recorded in some historical PMU data sets.

A general way to eliminate missing samples in a dataset is by simply deleting the missing values but this will reduce the total number of samples. There are other general methods like replacing the missing values by mean of the dataset, or mode of the dataset, etc. These methods seem to generate unacceptable results for real-time control as they do not involve prediction from the past data samples. Methods like zero order hold (ZOH) and first order hold (FOH) [6], [7] use linear prediction of the missing samples by extrapolating recent available samples. Another method called Lagrange interpolation method [4] is also based on linear prediction from past samples and has been discussed in this paper.

This paper focuses power flow simulations performed to measure robustness of the one shot control scheme used in [2] assuming 0-10% of missing PMU's data. Section II explains about response-basedone shot control scheme used in [2] in detail.

II. RESPONSE BASED SYSTEM & ONE SHOT CONTROLS

In early days, the transient stability prediction and control methodwas not response based. It was event based in which, for example, operators had to check the status of circuit breaker to get an idea of a fault in a particular area. Event based method essentially relies on direct detection of equipment outage. Response based system, however, involves real-time monitoring and analysis of continuous electrical phasors collected to a central location from different parts of the network via PMUs.

Different machine learning algorithms have been studied and used so far for transient stability prediction and/or control. Reference [9] is based on real time transient stability prediction using decision trees (DTs). Reference [10],[11] extends the method of using decision trees to trigger one-shot stabilizing control. In [1], the method of using DTs and wide area one shot controls are studied in a 176-bus model of Western Electrical Coordinating Council (WECC).

Subsections below discuss more about the power system model and methodology for one shot controls used in [1], [2] which are also the same methodologies used for this paper.

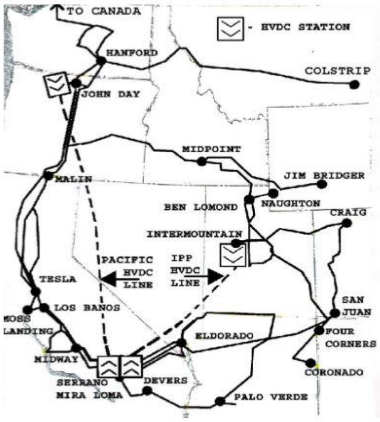

Fig.1. Transmission lines for the 176-bus model

Manuscript published on 30 December 2019.
***** Correspondence Author (s)
Niraj Dahal*, Steven M. Rovnyak, ECE Department, IUPUI, Indianapolis, USA. Email: ndahal@iu.edu
Based up on the work supported by NSF Award ECCS-1711521

A. Power System Model

A 176-bus model from WECC consisting of 29 generators and 71 load buses have been considered for this research. PMUs are placed at 17 buses.

Fig. 1 shows the transmission line diagram for the same model.

B. Phasor Measurements

Phasor measurements like bus voltage magnitudes, bus voltage angles, generator rotor angles, etc.from 17 different PMUs and 29 machines are collected at 30 Hz. The raw bus voltage angles wrap around -180 to +180 degrees and needs to be reconstructed as explained in [12]. Fig. 2 and Fig 3 shows a 12 second simulation of a contingency to show how wrapped bus voltage angles are reconstructed for continuous trajectories.

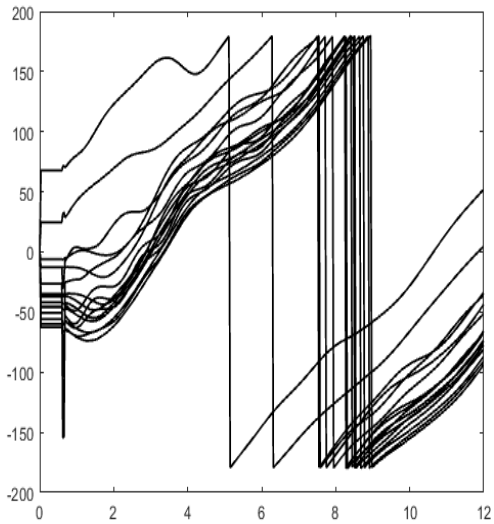

Fig.2. Wrapped bus voltage angles

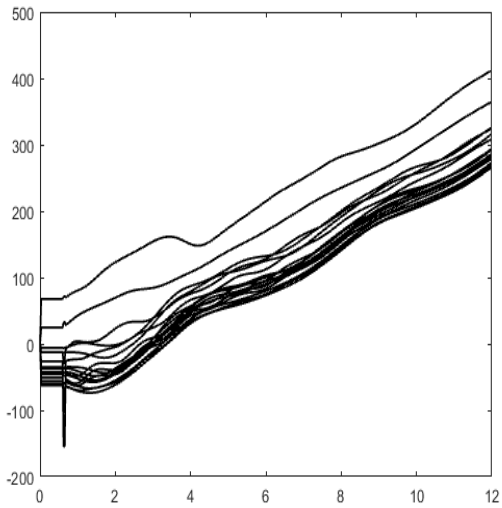

Fig.3. Reconstructed/unwrapped bus angles

C. One Shot Control Combination

One shot control is a single combination of controls that possibly reduce the phase angle differences when applied for preventing loss of synchronism. The one-shot control used in this research is based on real power injections at AC buses.

The control combination used here comprises 500 MW fast power increase in buses to MONTANA and CANADA and 500 MW fast power decrease to two buses near the southern end of the model. This control combination has been labelled as CC1 in this paper.

D. Training and Test Sets for DT

The training set contains 385 contingencies out of which 160 are single outages, 210 are double outages from WECC model and remaining 15 are single contingency events on the Pacific AC Intertie (PACI).

The test set used in this paper contains 480 single line to ground faults ranging from 4-15 cycles of fault duration. All the fault clearance are assumed to be at 40th cycle or 0.67 seconds.

E. Training DTs

The training is done using *fitctree()* function in MATLAB where input vector is composed ofbus voltage magnitudes and angles from 17 PMUs and indices [2], [13] calculated from those phasors. The DT targetis set to 'Control' or 'No Control' at every time stamp based on whether the simulation is unstable or stable respectively. The stability of any contingency is confirmed if the maximum rotor angles between any two generators at any time stamp does not exceed 300 degrees, otherwise it is unstable [12].

The misclassification cost during training is set 10 or 100 times greater for misclassifying 'No Control' compared to misclassifying 'Control' samples. The use of such higher relative misclassification cost is to decrease the number of control actions. Changing the misclassification cost can increase or decrease the number of events controlled.

Similarly, the complexity cost or maximum tree splits (*MaxNumSplits)* during the training can be used to control the size of the DT generated. This hyper parameter is set to 1 for generating a single node DT. The DT trained using bus angle velocity (Θ_{dot}) of 17 PMU buses classifies an event to be controlled when the bus angle velocity of 9th PMU bus equals or exceeds round-off and adjusted value of 50 degrees i.e. $(\Theta_{dot9} \geq 50)$.

In addition to the control DT, this paper also uses another DT to detect end of event. It is found that instead of checking the control DT continuously for every time-stamp, the success rate can be improved if the control DT is applied within a few seconds after detection of an event.

The following subsection discuss about training the event detection DT and its combination with the control DT for response-based control.

F. Training DTs for Event Detection

It is already mentioned that all the contingencies have short circuit to ground fault cleared at 0.67 seconds. The target for event detection DT is labelled as 'Postfault' for every time sample between 0.67 seconds to 0.72 seconds. For all other time stamps, it is 'Not Postfault'. The predictor for event detection DT is an index calculated as derivative of variance of bus voltage magnitudes (*bmvardot*) discussed in [2], [13], [14].

With the relative misclassification cost set to 1:100 and maximum splits set to 1, the DT gives 'Postfault' for any sample that has *bmvardot <= -0.03*.

Fig.4. shows the methodology for one shot control actuation using both event detection and Control DTs.To help reduce the number of unnecessary control actuations, the control DT is checked only for 0.167 seconds or 5 cycles after detection of an event.

Further, if the DT orders control, the one-shot control is applied 100 milliseconds (not shown in the flowchart) later in the simulation.

G. Performance Evaluation of One-Shot Control

The simulation measures control performances using the following parameters.

i) Stabilized simulation
ii) Destabilized simulation
iii) Simulation kept stable
iv) Simulation kept unstable
v) Simulation controlled
vi) Unnecessary Controls
vii) Average control time
viii) Total Success rate

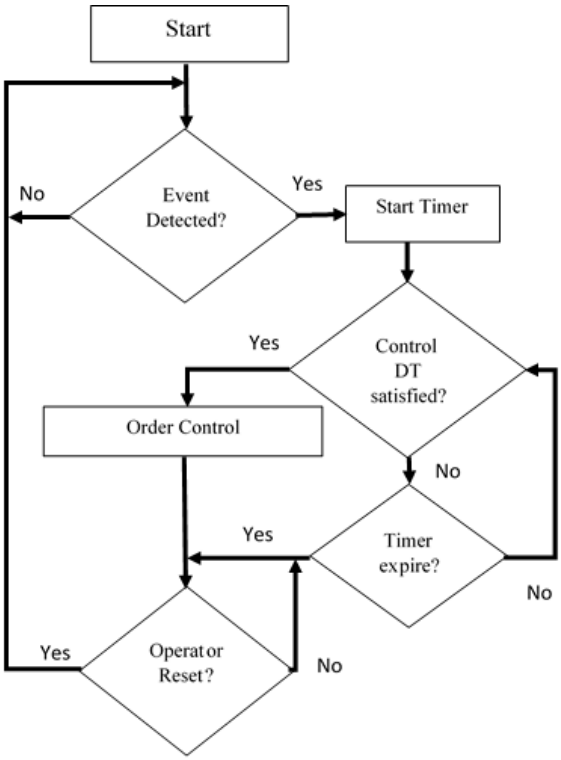

Fig. 4. Flow chart for control with event detection

The number of controlled simulation counts every simulation during which control is ordered and actuated. The number of stabilized simulation includes simulation made stable by the application of control. Destabilized simulation includes number of simulations made unstable by the application of control. Unnecessary controls are the simulations which are already stable but still actuates control and remain stable. Success rate is the ratio of number of stabilized simulation to the number of controlled simulations.

The average control time is the sum of individual control times of the events divided by total number of controlled simulations.

III. METHODS FOR HANDLING MISSING DATA

The paper focuses missing data recovery methods based on linear predictive model. A simple linear prediction or extrapolation is zero order hold (ZOH) which takes past single sample and keeps on holding it until the arrival of next sample. First order hold (FOH) uses weighted sum of two recent samples from the same channel to predict the missing value. We have also employed Lagrange polynomial method and use up to three previous samples for missing data imputation.

A. Data Hold

By holding a sample, a continuous time signal $g(t)$can be generated from a discrete-time sequence $x(kT)$ [7]. Considering a signal $g(t)$ during a time interval t where $kT \le t \le (k+1)T$, then $g(t)$ can be approximated by a polynomial as-

$$g(kT+\tau) = a_0 + a_1\tau + \cdots + a_{n-1}\tau^{n-1} + a_n\tau^n$$
$$g(kT+\tau) = x(kT) + a_1\tau + \cdots + a_{n-1}\tau^{n-1} + a_n\tau^n \quad (1)$$

where $0 \le \tau \le T$.At$\tau = 0$, $g(kT) = x(kT)$.

An nth order hold circuit uses $(n+1)$ discrete samples to generate $g(kT+\tau)$. In (1), if n=0, ZOH is obtained as in (2).

$$g(kT+\tau) = x(kT) \quad (2)$$

Fig. 5 shows input and output of azero-order hold circuit.

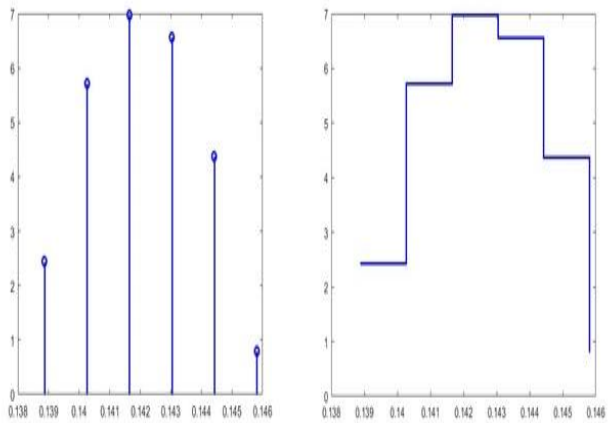

Fig. 5: Input (left) and Output (right) to ZOH

Similarly, a FOH is obtained from (1) when n=1 such that,

$$g(kT+\tau) = x(kT) + a_1\tau \quad (3)$$

We also have,

$$g\big((k-1)T\big) = x\big((k-1)T\big) \quad (4)$$

Therefore,

$$x(kT) - a_1T = x\big((k-1)T\big)$$

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

$$a_1 = \frac{x(kT) - x\big((k-1)T\big)}{T} \quad (5)$$

This implies,

$$g(kT+\tau) = x(kT) + \frac{x(kT) - x\big((k-1)T\big)}{T}\tau \quad (6)$$

For the same input samples as in Fig. 5, the first order hold would generate the graph in Fig. 6.

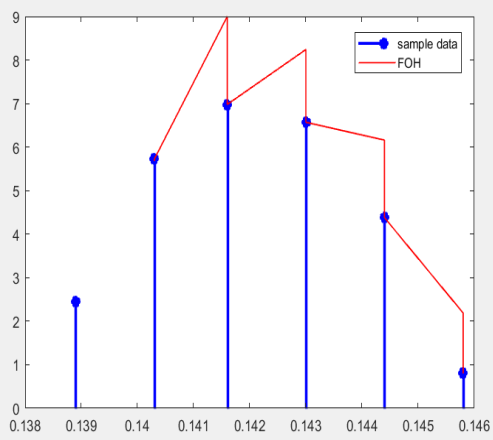

Fig. 6. Input and Output from FOH

B. Lagrange Polynomial Method

Lagrange interpolating polynomial is described in [4], [15], [16] and can be represented as follows-

$$L(x) = \sum_{j=1}^{n} y_j l_j(x) = \sum_{j=1}^{n} y_j \prod_{k=1,k\neq j}^{n} \frac{x - x_k}{x_j - x_k} \quad (7)$$

Here $x_j \neq x_k$ and $l_j(x)$ is the coefficient in the Lagrange polynomial.

Lagrange polynomial can be reduced and expanded for *n=3* as follows-

$$L(x) = y1\,\frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} + y_2\,\frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} + y_3\,\frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)} \quad (8)$$

Reference [16] classifies Lagrange polynomial method into different orders.

First Order Lagrange Polynomial (FO-Lag)

FO-Lag involves polynomial approximation considering two samples, one each from succeeding (S) and preceding (P) positions. The missing sample, denoted by M, can then be approximated from the polynomial.

Table I: FO-Lag Method

Preceding data (P)	Missing data (M)	Succeeding data (S)
x-1	x	x+1
y-1	y	y+1

In the table, all the x's are time domain samples and all the y's are phasor measurements.

Second Order Lagrange Polynomial (SO-Lag)

SO-Lag recovers missing sample considering two previous samples and one future sample as shown in Table II.

Table II: SO-Lag Method

P2	P1	M	S
x-2	x-1	x	x+1
y-2	y-1	y	y+1

Third Order Lagrange Polynomial (TO-Lag)

TO-Lag recovers missing sample considering three previous samples and one future sampleas in Table III.

Table III: TO-Lag Method

P3	P2	P1	M	S
x-3	x-2	x-1	x	x+1
y-3	y-2	y-1	y	y+1

The following method as illustrated in Table IV only uses preceding samples and hence is called strict Lagrange extrapolation (SE-Lag). This is one of the methods used in [5] for recovering missing value.

Table IV: SE-Lag Method

P3	P2	P1	M
x-3	x-2	x-1	x
y-3	y-2	y-1	y

The Lagrange coefficient $l_j(x)$ for each of the above methods can be determined as follows. As the phasor measurements are obtained at a sampling frequency of 30 Hz, the difference $(x_j - x_k)$ for two consecutive samples is constant and equal to (1/30). Table V provides the Lagrange coefficients calculated for the simulation.

Table V: Lagrange Coefficients

Methods	l_1	l_2	l_3	l_4
FO-Lag	1/2	1/2	-	-
SO-Lag	-1/3	1	1/3	-
TO-Lag	1/4	-1	3/2	1/4
SE-Lag	1	-3	3	-

IV. RESULTS AND ANALYSIS

The percentage of missing data from 0-10% is simulated by generating uniform pseudorandom numbers using *rand()*in MATLAB [17].

A. Performance using ZOH

Recovering missing samples by ZOH was expected to give lower performance, however, simulation results reflected an increase in the total number of events controlled as seen from Fig. 7. We assume that this behavior of ZOH to control more events is because of the fact that the difference between two samples at the recovery is nearly zero followed by a difference twice as large as it should be.

Retrieval Number: A10621291S319/2019©BEIESP
DOI:10.35940/ijeat.A1062.1291S319
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The large difference following the recovery causes the derivative to be large and hence forces the decision tree to actuate control for more number of simulations.Fig. 8 shows the number of events stabilized up to 10% missing data using ZOH.

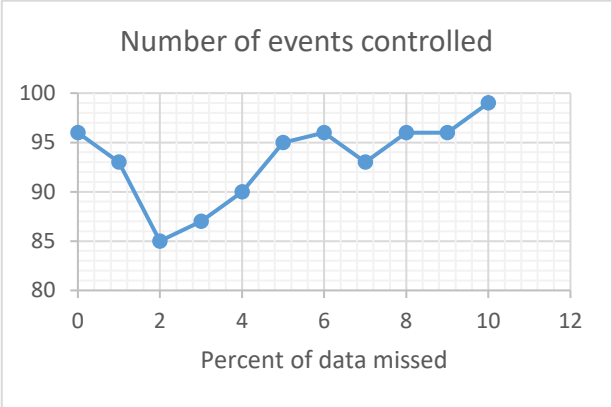

Fig.7: Events controlled using ZOH

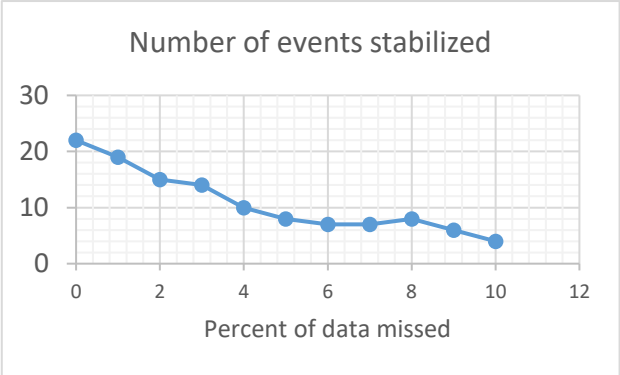

Fig. 8: Events stabilized using ZOH

B. Performance using FOH

FOH method seems relatively robust in terms of lowering the number of unnecessary controls. For the same percentage of missing samples, FOH has less number of unnecessary controls as shown in Fig. 9.

Fig. 10 shows the success rates for ZOH & FOH. Having a higher success rate refers that most of the simulations controlled resulted in stabilizing events hence is better. Trading off with either more stabilized events or less unnecessary controls, we can either use ZOH or FOH.

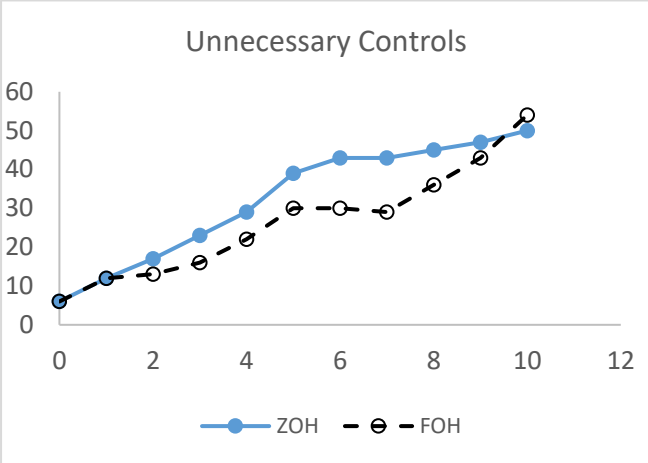

Fig. 9: Comparison of unnecessary controls vs % miss

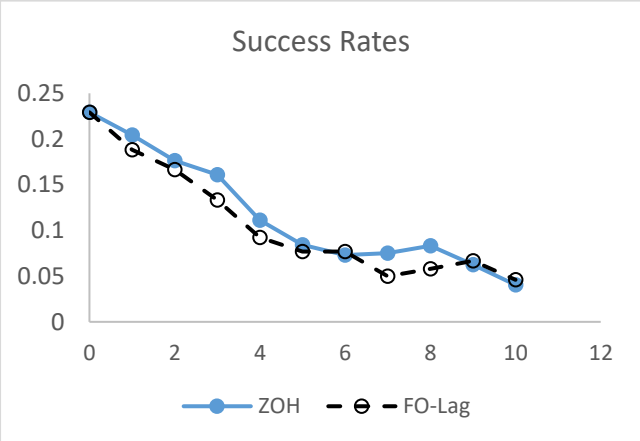

Fig. 10: Comparison of success rates vs % miss

C. Performance using Lagrange Polynomial Method

Fig. 11 compares the number of events stabilized by data hold methods and SE-Lag method. It is seen that the SE-Lag in average stabilizes lesser number of events than ZOH but more events than the FOH method.

Referring to Fig. 12, methods like FO-Lag, SO-Lag and TO-Lag prove exceptionally better than any other methods, but they cannot be used for real time missing data recovery. It is because these methods use one succeeding data point which would rather become a future value and unrealistic for real time recovery.

For all methods let's analyze the change in average control time (*Tavg*) for 480 SLG events for every percentage of data missed from 0-10%. Previous work in [2] mentions that the average control time for 480 SLG events with no data missing is 0.76 seconds. For each method, we now calculate standard deviation to measure the extent by which *Tavg* varies with increase in missing data.

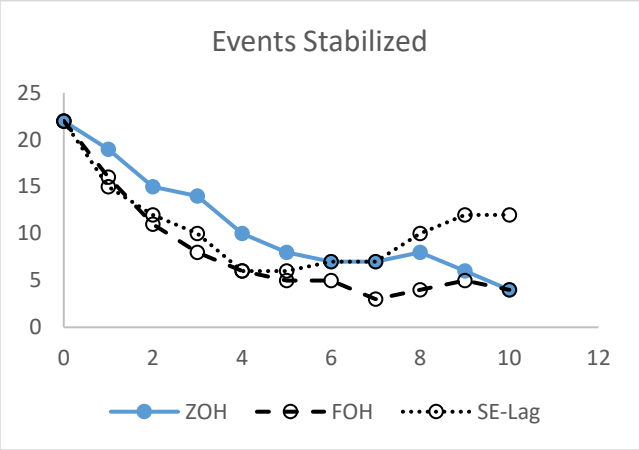

Fig. 11: Comparison of events stabilized vs % miss

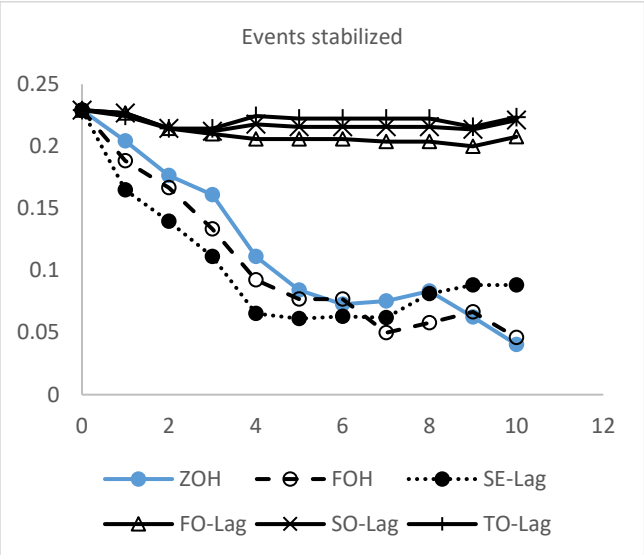

Fig. 12: Events stabilized for different methods

Standard Deviation can be calculated as-

$$\sigma = \sqrt{\sum_{i=1}^{N} \frac{\left(t - t_{avg}\right)^2}{N-1}} \qquad (9)$$

Here, t_{avg} = 0.76 seconds

t= observed value of T_{avg} from simulation

N = number of data = 11 (from 0-10%)

The standard deviation of T_{avg} for each method are found as follows:

i. ZOH = 0.329
ii. FOH = 0.419
iii. SE-Lag = 0.438
iv. FO-Lag = 0.026
v. SO-Lag = 0.023
vi. TO-Lag = 0.026

Even though the standard deviation for FO-lag, SO-Lag and TO-Lag methods are much lower, their approach to recover missing values in non-real time leads to a conclusion that ZOH has relatively superior performance.

V. CONCLUSION

Analyzing data retrieved from DSA tools like TSAT [18] shows a gradual change in bus angles and hence a simple extrapolation methods like ZOH and FOH are found better for missing data recoveryin a response-based control scheme as in [1], [2].

The first three order of Lagrange interpolating polynomial methods yielded much improved control performances, however, as these methods required the controller to use future values, the simulation of interpolation methods in this paper were not realistic for real time recovery as also stated in [19].

As a future enhancement, simulations can be performed to see how robust the one shot control is in a realistic noisy environment. Researchers can also evaluate the response based one shot control methods in a power system with different operating conditions or increased load condition.

REFERENCES

1. Q. Gao and S.M. Rovnyak, "*Decision trees using synchronized phasor measurements for wide-area response-based control*," IEEE Transactions on Power System, vol. 26, May 2011.
2. S. Rovnyak, D. W. Longbottom, D. C. Vasquez, and M. N. Nilchi, *Response based event detection for one shot wide-area stability controls. Chapter 8*: IET, 2017, ISBN:978-1-78561-477-4/ eBook: 978-1-78561-478-1.
3. J. D. L. Ree, V. Centeno, J. Thorp, and A. Phadke, "*Synchronized phasor measurement application in power systems*," IEEE Transactions on Smart Grid, vol. 1, no. 1, pp. 20-27, June 2010.
4. C. Huang, F. Li, and E. L. Zhan, "*Data quality issues for synchrophasor applications part II: problem formulation and potential solutions*," J. Mod. Power System Clean Energy. 2016.
5. M. Wang, "*Data quality management of synchrophasor data in power system by exploiting low-dimensional models*," 51st Annual Conference on Information Sciences and Systems (CISS), Baltimore, MD, pp. 1-2, 2017.
6. Z. Zhang and K. T. Chong, "*Comparison between first order hold with zero order hold in discretization of input-delay nonlinear systems*." International Conference on Control, Automation and Systems, Seoul, pp. 2892-2896, 2007.
7. Data hold methods, [Online]. Available:http://web.cecs.pdx.edu/tymerski/ece452/Chapter3.pdf
8. C. Taylor, J. Haner, L. Hill, W. Mittelstadt, and R. Cresap, "*A new out-of-step relay with rate of change of apparent resistance augmentation*," IEEE Transactions on Power Apparatus and System, vol. PAS-102, no. 3, pp. 631-639, March 1983.
9. S. Rovnyak, S. Kretsinger, J. Thorp, and D. Brown, "*Decision trees for real-time transient stability prediction*," IEEE Transactions on Power System, vol. 9, no. 3, pp. 1417-1426, August 1994.
10. K. Mei and S. M. Rovnyak, "*Response based decision trees to trigger one shot stabilizing control*,". IEEE Transactions on Power System, vol 19, pp. 531-537, Feb. 2004.
11. S. M. Rovnyak, G. Li, and k. Mei, "*One-Shot controls for preventing loss of synchronism*," IEEE Power Engineering Society General meeting, 2003.
12. M. Nilchi, "*Electric utility planning methods for the design of one-shot stability controls*," MS Thesis, Purdue University, Indianapolis, 2012.
13. S. Rovnyak, M. Nilchi, D. Longbottom, and D. Vasquez, "*Angle stability predictive indices*," IEEE Power and Energy Society General Meeting, San Diego, CA, pp. 1-6, 2012.
14. G. Li and S. M. Rovnyak, "*Integral square generator angle index for stability ranking and control*," IEEE Transactions on Power System, vol. 20, May 2005.
15. N. Le and W. Benjapolakul, "*A data imputation model in phasor measurement units on bagged averaging of multiple linear regression*," IEEE Access, vol. 6, pp. 39324-39333, 2018.
16. P. Manembu, A. Kewo, and B. Welang, "*Missing data solution of electricity consumption based on lagrange interpolation case study: intelligentsia data monitoring*," 2015 International Conference on Electrical Engineering and Informatics (ICEEI), Denpasar, pp. 511-516, 2015.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

17. MATLAB, [Online] Available: https://www.mathworks.com/products/matlab.html
18. Powertech labs, inc. TSAT: Transient security assessment tool, [Online]. Available: https://www.dsatools.com/wp content/uploads/2017/07/TSAT.pdf
19. N. Dahal, "*Methods of handling missing data in one shot response-based power system control*," MSECE Thesis, Purdue University, Indianapolis, 2019.

AUTHORS PROFILE

Niraj Dahalis currently a PhD student in electrical engineeringat Indiana University, Purdue University, Indianapolis (IUPUI) specialized in power and controls. He completed master's in electrical engineering from the same university in 2019. He has been actively involved in research about pattern recognition for one shot controls in power system. His area of interest includes transient stability prediction and controls, small signal stabilities, and use of machine learning algorithms to ensure dynamic security in power system. He has been a student member of IEEE and IEEE Power and Energy Society.

Steven M. Rovnyak was born in Lafayette, IN, on July 4, 1966. He received the A.B. degree in mathematics in 1988 and the B.S., M.S., and Ph.D. degrees in electrical engineering in 1988, 1990, and 1994, respectively from Cornell University, Ithaca, NY. Currently, he is Associate Professor of Electrical and Computer Engineering at Indiana University - Purdue University Indianapolis (IUPUI). Previously, he was an Assistant Professor of Electrical Engineering at Louisiana Tech University, Ruston, and spent two years as Postdoctoral Associate at Cornell University, Ithaca, NY. He worked on Electric Power Research Institute (EPRI) projects while a graduate student and Postdoctoral Associate at Cornell. He has developed algorithms to process wide-area monitoring system (WAMS) data with grant support from Entergy Transmission. His research interests include the development of pattern recognition methodologies for protective relaying and for one-shot stability controls. Dr. Rovnyak is a member of the Phi Beta Kappa and Phi Kappa Phi honor societies.