

# PMU Time Error Detection Using Second-Order Phase Angle Derivative Measurements

Ikponmwo Idehen and Thomas J. Overbye

**Abstract**— Significant loss or corruption of time and message control bit information in reported PMU measurements can lead to a loss of data quality awareness by engineers and control center applications. In this paper, four different time-based issues encountered in PMU operation, and corresponding, unique data error propagation models in measurements of second derivative phasor angles are discussed. As part of a data post-processing step, a novel application of subsequence dynamic time warping technique is proposed for use in a similarity-based, time-series, pattern query to identify instances of PMU internal clock drift time error. Measurements from a synthetic 2,000-bus network have been used to test the effectiveness of the proposed method.

**Index Terms**— phasor measurement unit, time-series, error analysis, pattern recognition, time-synchronization

## I. INTRODUCTION

TIME synchronization of phasor measurements is a unique technological feature utilized by phasor measurement units (PMUs) and other synchrophasor devices, and largely accounts for their widespread deployment in grid monitoring [1-3]. Unfortunately, in the event of device failures, this unique operation mode renders a new paradigm of measurement errors, and expands data quality issues tackled by control centers. PMUs require global positioning system (GPS) signal receivers to synchronize internal sampling clocks to a reference signal from a Coordinated Universal Time (UTC) [4]. While normal device operation generates precise and time-synchronized phasor measurements, component malfunctions often translate to inaccurate measurements. An expanded suite of data integrity checks is thus needed to assure high accuracy and quality of PMU measurements used in control center data analytics.

Several time-based PMU device errors have been reported in the literature [6-12]. Generally associated with internal clock malfunctioning, partial or full GPS signal loss, and other cybersecurity issues, these multiple time signal delivery failure modes manifest as biases in phasor angle measurements. Depending on the scale and duration of anomaly in the PMU infrastructure, significant deviations in phasor measurements are reported which can result in false conclusions about the state of the grid when these measurements are utilized by critical grid monitoring applications [13]. In order to aid identification of time errors, [14] has incorporated time and message quality flag bits in PMU IEEE C37.118.2 data - binary, flagged bits, used as control data, provide time synchronization status and quality of PMU data information. However, [15-17] describe how synchrophasor measurements could be rendered unintelligible through due to a loss of logical dataset consistency bits mislabeling, duplication, wrong device identifiers or erroneous time stamping occur in recording

devices. In addition, control bit data loss or corruption, due to compression and data transformations in the PMU data archival process were also noted to contribute to undecipherable measurements. An overall impact is a loss of awareness when engineers and control center applications depend on these unreliable datasets for grid visibility. A contingency plan for data quality check is thus required in the event of loss of critical message and time quality control bits in reported PMU data.

Studies in PMU error mechanisms reveal a non-dependence of the magnitude component of synchrophasor measurements to the effect of time-based errors [7-9]. These time errors manifest in phasor angle measurements, and their first and second derivatives - frequency and rate of change of frequency (ROCOF) measurements. In this paper, an expatiation of the uncommon and unique patterns of ROCOF data measurements, due to selected PMU time error mechanisms, is presented. Given different data signatures, a similarity-based, time series pattern query is proposed for the search of instances of time-related errors in strings of reported PMU measurements. The contribution of this paper is to provide an alternative identification technique, as part of a data post-processing stage, for detecting time errors in PMU measurements. It leverages developed error propagation models to generate prototypes of PMU error signatures.

## II. CHARACTERIZATION AND ERROR-PROPAGATION OF TIME-BASED PMU DATA ERRORS

In this section, four different PMU time errors are discussed using respective phase angle and ROCOF time-series measurements, followed by the formulation of the corresponding time error propagation models.

### A. Time Errors and Error Propagation Models

Assume an off-nominal system frequency of  $f_i$  Hz, the phase angle deviation  $\Delta\delta_\epsilon$  due to a time error  $\Delta t_\epsilon$ , can be computed from (1).

$$\Delta\delta_\epsilon = 360\Delta t_\epsilon f_i \quad (1)$$

where  $f_i = f_o + \Delta\epsilon$ .  $f_o$  and  $\Delta\epsilon$  are the nominal frequency and frequency deviations respectively. A component of  $\Delta\delta_\epsilon$  due to  $\Delta\epsilon$  is  $360\Delta t_\epsilon \Delta\epsilon$ , such that  $\Delta t_\epsilon$  is in the order of microseconds. Neglecting the effect of  $\Delta\epsilon$  (since  $\Delta\epsilon \in (0,0.05)$ ) in normal operating conditions,  $\Delta\delta_\epsilon = 360\Delta t_\epsilon f_o$

Phase angle error due to the observed time difference,  $\Delta t_\epsilon$  for any reported measurement, however is dependent on the time error source. A generalized expression for computing the instantaneous phase angle error at any reported time from device error inception is shown in (2),

$$\Delta\delta_{\varepsilon,t}(t) = 360\Delta t_{\varepsilon}(t)f_o \quad (2)$$

Figs. 1 and 2 illustrate voltage angle (VA) profiles observed in four different time error prototypes during a 30 second simulation carried out in a PMU error emulation environment [18]. Error injections are carried out at a test location using a 2,000-bus synthetic network [19], while grid frequency and per-second device report rate are set to 60 Hz and 30 samples respectively. Based on a chosen time error, phasor magnitude and angle measurements are modified to new values, while frequency and ROCOF data are measured using the method of weighted least-squares (WLS) estimation technique - it assumes three-phase balanced inputs, such that only positive sequence data are needed [20]. The process of error injection is initiated at the 5<sup>th</sup> second, and exists for 20 seconds. Ideal measurements, devoid of noise signals, have been used in order to isolate the effects of time errors.

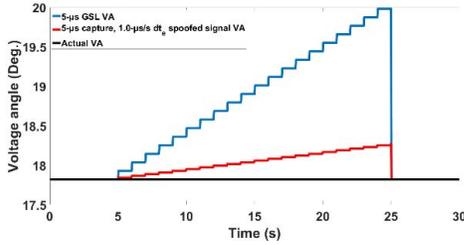


Fig. 1. Prototyped PMU voltage angles for GSL and time signal spoofing

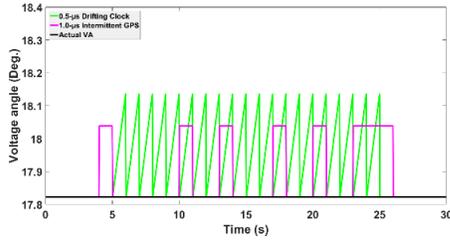


Fig. 2. Prototyped PMU voltage angles for internal clock offset and intermittent GPS reception

Fig. 1 illustrates the profile for a GPS signal loss (GSL) event and a spoofed-GPS time signal at a test bus location. A black, solid line is the true, steady state VA. The blue-colored GSL event is simulated while assuming a pulse per second (PPS) time error ( $\Delta t_{\varepsilon}$ ) of  $5 \mu s$ , and the red-colored spoofed signal event is given a time error divergence rate ( $dt_e$ ) of  $1 \mu s/s$ . Based on defined error mechanisms in [11, 12], time error propagation models deduced in this work for the GSL and spoofed signal events are given by (3) and (4) respectively.

$$\Delta t_{\varepsilon,t}(t) = t\Delta t_{\varepsilon} \quad (3)$$

$$\Delta t_{\varepsilon,t}(t) = \Delta t_{\varepsilon,t-1}(t-1) + t \times dt_e \quad (4)$$

where  $t$  is the reporting time (in seconds) upon error initiation,  $\Delta t_{\varepsilon}(t)$  is the instantaneous, accumulated time drift at time  $t$ ; and  $dt_e$  is the rate of time signal divergence induced by an attacker. In the spoofed signal event, the time error  $\Delta t_{\varepsilon}$  at an instance of complete capture of the GPS receiver (when  $t = 1$ ) is observed to include a time offset,  $\Delta t_{\varepsilon,capture}$  [12]. For each of the events, the phase angle error  $\Delta\delta_{\varepsilon}(t)$  is applied uniformly

on all 30 samples in a one-second reporting period prior to the next set of reported samples. Equations (3) and (4) indicate periodic time error accumulation in the phasor angle propagation models of both GSL and spoofed receiver time signal events until the completion of error injection. The incrementing GSL angle errors are illustrated by the observed deviations from the true VA measurement. However, angle errors for the spoofed signal are less observable, and only become visible long after the moment of receiver capture.

Fig. 2. illustrates VA profiles due to two different causes of PMU clock time offsets – a constant  $0.5 \mu s$  time error as a result of a drifting internal clock, and a  $1.0 \mu s$  error due to intermittent GPS clock signals received by the PMU device. Based on [8, 9], the deduced time error propagation models for both cases are given respectively as,

$$\Delta t_{\varepsilon,i}(t) = (i-1)\Delta t_{\varepsilon}, i = 1, 2 \dots n \quad (5)$$

$$\Delta t_{\varepsilon,t}(t) = \Delta t_{\varepsilon} \quad (6)$$

where  $n$  is the report rate of the device. A non-uniform time error at all  $n$  samples within a reporting interval is indicated by (5). The green-colored ramp of Fig. 2 is indicative of an increasing angle deviation at each sample when the device clock is observed to drift away. However, at the reception of a GPS re-synchronizing pulse (i.e. when  $i = 1$ ), a time reset for the next reporting cycle ensures the accuracy of the first sample. Since the internal clock has a constant time offset, the cycle repeats, thus the periodic VA ramps. In contrast, the pink-colored, intermittent GPS signal introduces a constant time error in the PMU clock which is then applied uniformly across all samples, and at any reporting time  $t$  of its occurrence. The result is an otherwise, non-periodic VA profile in Fig. 2.

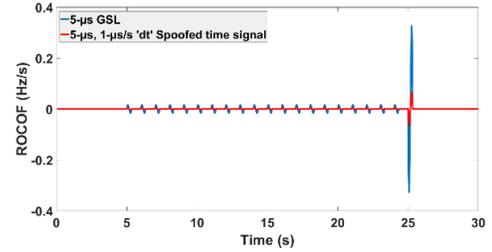


Fig. 3. ROCOF data for prototyped PMU voltage angles in Fig. (1)

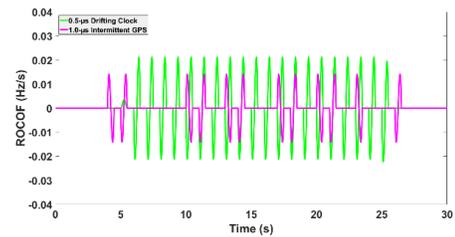


Fig. 4. ROCOF data for prototyped PMU voltage angles in Fig. (2)

PMUs report ROCOF data which can also be used to monitor phasor angle changes. ROCOF data for the discussed time errors are shown in Fig. 3 and 4. Periodic ripples observed in Fig. 3 for both GSL and spoofed GPS signal are attributed to the small jumps in voltage angles due to the incremental time errors. However, at the point of error removal, the accumulated

voltage angle deviation results in a sudden spike in the ROCOF. It is observed that the GSL event generates a significant spike ( $\sim 0.33$  Hz/sec) due to the large angle deviation as compared to the spoofed time signal.

Given unique patterns in phasor angle-derived measurements, and some of the quality issues posed to PMU control data mentioned in the previous section, an alternative time-error identification method in PMU data is proposed. The unique voltage angle and ROCOF patterns provide a basis for a similarity search for specific errors in PMU data. This work assumes prior detection of the grid location reporting bad data.

### III. THEORETICAL BACKGROUND

#### A. Similarity Matching- Illustrating Example

Given two data sequences:

$$\begin{aligned} \mathbf{X} &= \{1,1,1,3,3,3,3,3,3,1,1,1,1\} \\ \mathbf{Y} &= \{1,1,1,1,1,3,3,3,3,3,3,1,1,1\} \end{aligned}$$

A dis-similarity value ( $\rho$ ) between the sequences can be computed using the  $L_n$  norm [21, 22].

$$\rho = \sqrt[n]{\sum_i^m (x_i - y_i)^n} \quad (7)$$

$x_i \in \mathbf{X} \in R^m$  and  $y_i \in \mathbf{Y} \in R^m$

When  $n = 2$ , the computed dissimilarity value (Euclidean distance) is 5.66, and indicates a disparity between both sequences. Further observation reveals  $\mathbf{X}$  and  $\mathbf{Y}$  have exact element values however, shifted in time. The Euclidean measure fails to detect a time alignment issue, and produces a non-intuitive similarity value. Furthermore, given a pattern sequence,  $\mathbf{A} \in R^\alpha$ , and a longer data sequence  $\mathbf{B} \in R^\beta = \{n_1, \dots, n_\beta\}$  such that  $\beta > \alpha$ , the goal to determine the similarity level between  $\mathbf{A}$  and  $\mathbf{B}$  of mismatched lengths is not captured by this Euclidean distance. Hence, the need for a metric that is invariant to time shifting, and with the ability to handle data sequences of differing lengths.

#### B. Dynamic Time Warping (DTW)

DTW is a transformation that temporally wraps a data sequence along its time axis in order to detect similar sequences [23-25]. The technique can be used to find an optimal alignment between sequences  $\mathbf{X}$  and  $\mathbf{Y}$ , while simultaneously computing a similarity level, otherwise known as the *DTW* distance. Considering two sequences,  $\mathbf{C} = \{c_1, c_2 \dots c_M\}$  and  $\mathbf{Q} = \{q_1, q_2 \dots q_N\}$ , an application of DTW requires an understanding of important terms in the literature [23-25].

Firstly, a cost matrix,  $\mathbf{P} \in R^{N \times M}$  is a 2-dimensional matrix containing the pairwise local cost between data points in  $\mathbf{C}$  and  $\mathbf{Q}$ . The goal is to find an alignment between  $\mathbf{C}$  and  $\mathbf{Q}$  with a minimal overall cost. In this paper, we have chosen the local cost to be the pairwise distance between data points i.e.  $P(i, j) = d(c_i, q_j)$  such that  $1 \leq i \leq N$  and  $1 \leq j \leq M$ .

Secondly, the warping path is a sequence  $p = (p_1, \dots, p_L)$  with  $p_l = (n_l, m_l) \in [1:N] \times [1:M]$  for  $l \in [1:L]$ .  $p_l$  can be described as a node in  $p$ , which in turn defines an alignment between  $\mathbf{C}$  and  $\mathbf{Q}$  subject to the following conditions:

a) *Boundary condition*:  $p_1 = (1,1)$  and  $p_L = (N, M)$ .

This is similar to a stretch on the time axis, thus ensuring that terminals of both sequences are aligned to each other.

b) *Monotonicity condition*:  $n_1 \leq n_2 \leq \dots \leq n_L$  and  $m_1 \leq m_2 \leq \dots \leq m_L$ . This condition forces the different nodes to be monotonically spaced in time.

c) *Step size condition*:  $p_{l+1} - p_l \in \{(1,0), (0,1), (1,1)\}$  for  $l \in [1:L-1]$ . This constraint ensures that every element in  $\mathbf{C}$  and  $\mathbf{Q}$  is used in the formation of all possible  $p$ 's in  $\mathbf{P}$ .

Thirdly, the total cost,  $d_p(\mathbf{C}, \mathbf{Q})$  of a warping path,  $p$  is the sum of all local cost measures,  $d$  in  $p$ . i.e.  $d_p(\mathbf{C}, \mathbf{Q}) = \sum_{l=1}^L (dp_l)$  where  $dp_l = d(c_{n_l}, q_{m_l})$ . This applies to the summation of all local costs in any warping path,  $p$ , and is known as the *DTW* distance.

Finally, the optimal warping path,  $p^*$  is that path which minimizes the *DTW* distance among all the different possible paths.  $p^*$  is computed as  $\text{argmin}(d_p(\mathbf{C}, \mathbf{Q}))$ , and with a *DTW* distance of  $d_{p^*}(\mathbf{C}, \mathbf{Q})$ .

An alignment of two data sequences, and determination of an optimal *DTW* distance requires the construction of an accumulated cost matrix,  $\mathbf{AP} \in R^{N \times M}$  in a forward direction, followed by the determination of  $p^*$  in a backward direction. Both steps involve recursive calculations which rely on the use of dynamic programming (DP) algorithms, and are stated thus,

##### 1) Forward Direction

$$\begin{aligned} \mathbf{AP}(i, j) &= \mathbf{P}(i, j) + \min \begin{cases} \mathbf{AP}(i, j-1) \\ \mathbf{AP}(i-1, j) \\ \mathbf{AP}(i-1, j-1) \end{cases} \quad (8) \\ \mathbf{AP}(0,0) &= 0, \mathbf{AP}(i,0) = \mathbf{AP}(0,j) = \infty \\ &\quad (1 \leq i \leq M, 1 \leq j \leq N) \end{aligned}$$

Hence,  $(\mathbf{C}, \mathbf{Q}) = \mathbf{AP}(M, N)$ .

Initialization of the entries in the first row and column is carried out by an additional  $0^{\text{th}}$  row and column with values set to arbitrarily large numbers.

##### 2) Backward Direction (OptimalWarpingPath)

Beginning at  $p_l = (M, N)$

Determine:  $p_{l-1} :=$

$$\begin{cases} (1, m-1), & \text{if } n = 1 \\ (n-1, 1), & \text{if } m = 1 \\ \text{argmin}\{\mathbf{AP}(n-1, m-1), \mathbf{AP}(n-1, m), \mathbf{AP}(n, m-1)\}, & \text{otherwise} \end{cases} \quad (9)$$

$m \in [1:M], n \in [1:N]$

Computation of  $p^*$  involves starting at the  $(M, N)^{\text{th}}$  node and recursively tracing back the  $\mathbf{AP}$  matrix through the path of least cost until  $p_1$  is reached. An optimal path lying along the diagonal position of the matrix indicates a strong point-to-point correlation of the data points in both sequences.

Fig. 5 illustrates the  $\mathbf{AP}$  matrix for  $\mathbf{X}$  and  $\mathbf{Y}$  sequences, where  $N = M = 15$ . The matrix is read from left-to-right, and top-to-bottom for  $\mathbf{Y}$  and  $\mathbf{X}$  respectively. Node  $(M, N)$  is located at the right-bottom corner of the matrix, and with a *DTW* distance of 0. The heavy shaded area is the region of least cost, while the lightly shaded area is the optimal warping path. Initializing conditions for the matrix have been set to 999 i.e.  $\mathbf{AP}(i, 0) = \mathbf{AP}(0, j) = 999$ .

*Modification*: Multiple low cost nodes observed in different paths indicate exact correlation between values in  $\mathbf{X}$  and  $\mathbf{Y}$ . At any node,  $p_i$  the choice of  $p_{i-1}$  in the backward direction algorithm requires that it is the lowest cost node. In

this paper, an algorithm bias is introduced to ensure that in the event of multiple nodes with equal and minimal costs, the diagonal node is always preferred.

$$p_{i-1} = \underset{\text{argmin}}{\text{AP}}(n-1, m-1), \text{AP}(n-1, m), \text{AP}(n, m-1) \\ \text{if } \{p_{i-1}\} > 1, \text{ and } (n-1, m-1) \in p_{i-1} \\ \text{Return: node } (n-1, m-1) \quad (10)$$

	Y	1	1	1	1	3	3	3	3	3	3	3	1	1	1	
Index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
X	0	999	999	999	999	999	999	999	999	999	999	999	999	999	999	999
1	1	999	0	0	0	0	4	8	12	16	20	24	28	28	28	28
1	2	999	0	0	0	0	4	8	12	16	20	24	28	28	28	28
1	3	999	0	0	0	0	4	8	12	16	20	24	28	28	28	28
3	4	999	4	4	4	4	0	0	0	0	0	0	0	4	8	12
3	5	999	8	8	8	8	0	0	0	0	0	0	0	4	8	12
3	6	999	12	12	12	12	0	0	0	0	0	0	0	4	8	12
3	7	999	16	16	16	16	0	0	0	0	0	0	0	4	8	12
3	8	999	20	20	20	20	0	0	0	0	0	0	0	4	8	12
3	9	999	24	24	24	24	0	0	0	0	0	0	0	4	8	12
3	10	999	28	28	28	28	0	0	0	0	0	0	0	4	8	12
1	11	999	28	28	28	28	4	4	4	4	4	4	4	0	0	0
1	12	999	28	28	28	28	8	8	8	8	8	8	8	0	0	0
1	13	999	28	28	28	28	12	12	12	12	12	12	12	0	0	0
1	14	999	28	28	28	28	16	16	16	16	16	16	16	0	0	0
1	15	999	28	28	28	28	20	20	20	20	20	20	20	0	0	0

Fig.5. Accumulated cost matrix for  $X$  and  $Y$

### C. Subsequence DTW (SDTW)

A special form of DTW is to search for smaller, specific data sequences in long data streams. If  $N \gg M$ , we can search for a sub-sequence  $Q(a^*: b^*) := (q_{a^*}, q_{a^*+1} \dots q_{b^*})$  such that  $1 \leq a^* \leq b^* \leq N$ , and minimizes the DTW distance to  $C$  over all possible subsequences of  $Q$ . i.e.

$$(a^*, b^*) := \underset{(a,b): 1 \leq a \leq b \leq N}{\text{argmin}} (DTW(C, Q(a:b))) \quad (11)$$

A distinguishing feature in the modified algorithm is the relaxation of the previously-specified warping path boundary conditions.

$$\sum_{i=1}^m \mathbf{P}(i, 1) \text{ for } m \in [1:M] \text{ and } \mathbf{AP}(1, n) := \mathbf{P}(1, m) \\ \mathbf{AP}(i, 0) = \infty, i \leq i \leq M \\ \mathbf{AP}(0, j) = 0, i \leq j \leq N \quad (12)$$

Hence, the procedures to obtain  $b^*$ , and then  $a^* \in [1:M]$  are stated as follows:

- 1) Search for all  $b^*$ 's which minimize  $\mathbf{AP}(M, :)$  i.e.
$$b^* = \underset{b: [1:N]}{\text{argmin}} (\mathbf{AP}(M, b)) \quad (13)$$
- 2) To find  $a^*$ , begin backward recursive search from  $p_l = (M, b^*)$ . Apply the *OptimalWarpingPath* algorithm such that  $p_1 = (a^*, 1)$  for some  $l \in [1:L]$

## IV. CASE STUDY AND DISCUSSIONS

Motivated by the similarity-matching technique discussed in Section III, the SDTW method is applied in the search for specific time-based errors in PMU-sourced data.

### A. Prototype - Clock Delay

$\mathbf{P} \in R^M$  is a prototype pattern for a specific time-based error, and  $\mathbf{T} \in R^N$  is test ROCOF measurement obtained from a device.  $M$  and  $N$  are the number of data points in  $\mathbf{P}$  and  $\mathbf{T}$  respectively. Fig. 6 is an error prototype,  $\mathbf{P}_1$  used for the demonstration, and represents a  $5 \mu\text{s}$  PMU internal clock delay.

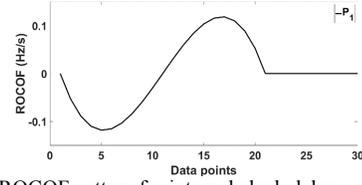


Fig.6. Prototype ROCOF pattern for internal clock delay

The time resolution used in the generation of this prototype error is 0.033 seconds.  $\mathbf{P}_1$  is at the instance of PMU clock re-synchronization with an external pulse per second (PPS), and has a duration of one reporting cycle comprising of 30 data points. The length of  $\mathbf{P}_1$  defines the  $M$ -dimension.

### B. Test Data

The algorithm was tested on three different erroneous test ROCOF datasets having a  $3 \mu\text{s}$  internal clock delay. Synthetic voltage angle measurements were obtained from the transient stability study of a 60 Hz, 2,000-bus synthetic network observed to show small variations in system loads. Using the error simulation platform in [20], time-based errors were injected into the voltage angles, from which ROCOF data stream was derived. A steady state, non-event data ( $\mathbf{T}_1$ ), dynamic state, event data ( $\mathbf{T}_2$ ), and a non-event, noisy data ( $\mathbf{T}_3$ ), each of duration of 10, 30 and 10 seconds respectively were simulated. Report rate of the PMU were set to 30 samples per second, such that  $N$  was set to 300, 900 and 300 data points for the different test cases respectively.

### C. Simulations

#### a) Steady state with small random load perturbations

Test ROCOF data  $\mathbf{T}_1$ , obtained from a test bus is shown in Fig. 7. The six measurement spikes correspond to an induced  $3 \mu\text{s}$  internal clock error which lasted for 6 seconds. In the absence of significant system dynamics and noise,  $\mathbf{T}_1$  is vividly observed to have a maximum of 0.07 Hz/s when clock delay as simulated between the second and eighth second of the total time.

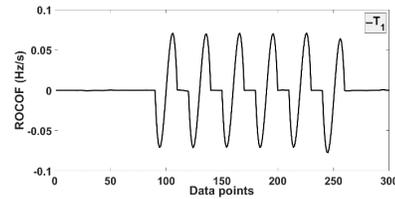


Fig. 7. 10-s non-event test ROCOF data  $\mathbf{T}_1$  for case study (a)

Fig. 8 is a cut-section of the accumulated cost matrix,  $\mathbf{AP} \in R^{30 \times 300}$  generated when the SDTW algorithm was run on  $\mathbf{T}_1$  in a search for  $\mathbf{P}_1$ . The directions for navigating through the matrix are similar to the illustration in Section III. The figure shows the first observed warping path traced as the yellow-colored sets of nodes, and observed between data points 90 and 110 in  $\mathbf{T}_1$ . Surrounding orange-colored cells indicate low-cost, alternative neighbor nodes. Since  $\mathbf{AP}$  is non-square, no exact diagonal exists for this path in order to capture identical relationship between  $\mathbf{P}_1$  and the first feature in  $\mathbf{T}_1$ . Different stacks of vertically and horizontally-aligned nodes within any optimal path capture a one-to-many correlation between data points in  $\mathbf{P}_1$  and  $\mathbf{T}_1$ . This is attributed to the time-stretch and data point mapping along different time points in both

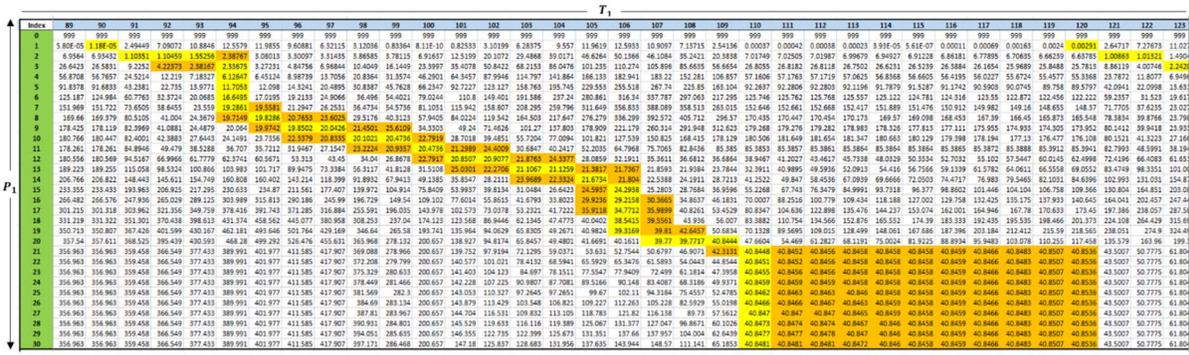


Fig. 8. Cut-section of accumulated cost matrix

measurements, which then uncovers the time- error pattern  $P_1$  in  $T_1$ . Incremental costs of alternative low-cost nodes are minimal or zero at worst, and a combination of nodes with those lying along the optimal warp path gives rise to redundant and incomplete sub-optimal paths of time-error patterns. These are similar to overlapping, neighbor subsequences in a time series.

Fig. 9 shows a plot of the  $DTW$  distance computed for every data point in  $T_1$ .

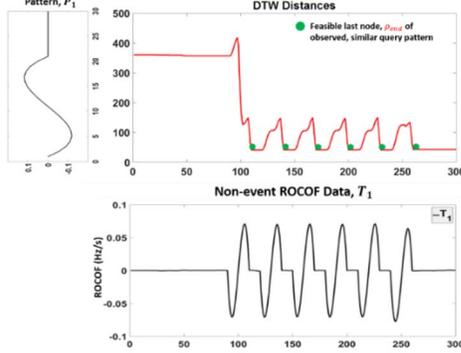


Fig. 9. Noise-free ( $P_1, T_1$ )  $DTW$  distances in  $T_1$

Visual inspection reveals an initially high  $DTW$  distance between  $P_1$  and the early data segments of  $T_1$  which is indicative of a large disparity between both signals. This value suddenly drops off ( $\sim 41$ ) once the first unique feature is completely identified in the  $T_1$  subsequence ( $SS_1 = T_{1:90}:T_{1:110}$ ) by a full cycle of  $P_1$ . This is represented by the first, green ball in the low-cost valley region of the graph. Successive rise-and-falls in  $DTW$  distances are detected in consecutive valley regions with similar  $DTW$  distances. These account for the remaining instances when  $P_1$  is identified in  $T_1$ .

A visual inspection and comparison of computed  $DTW$  values provides a firsthand identification of the presence of query, time error data patterns in any reported ROCOF test measurement. In addition, use of a defined threshold,  $DTW_{th}$  can aid pattern identification in the data, such that any data point in the test data with  $DTW$  distance smaller than  $DTW_{th}$ , is flagged as a data event instance. This parameter will depend on the maximum value of the query time error data pattern.

#### b) System event with line outage

A dynamic event is simulated with the outage of a 230-kV rated transmission line on the system, in addition to the internal clock delay at the test bus. The simulation was carried out for 30 sec.

Fig. 10 shows test bus measurement,  $T_2$ .

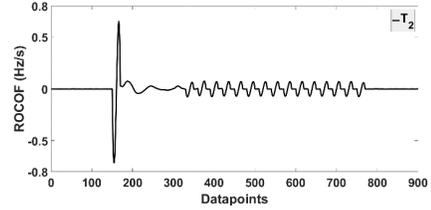


Fig. 10. 30-s event test ROCOF data  $T_2$  for case study (b)

The initial measurement spike ( $\sim 0.7$  Hz/s) is due to the line outage event occurring at the fifth second followed by those as a result of the induced clock delay error. These surges in ROCOF measurements obscure smaller variations which are actually due to system loading. Data points associated with the line outage event need to be removed prior to the execution of the similarity-based algorithm. For this paper, a variance-based event detection method [26] is applied to phasor voltage magnitude measurements to identify and replace eventful data points. Using the actual voltage measurements, event time points were identified, after which the corresponding ROCOF data points were replaced by a moving window average.

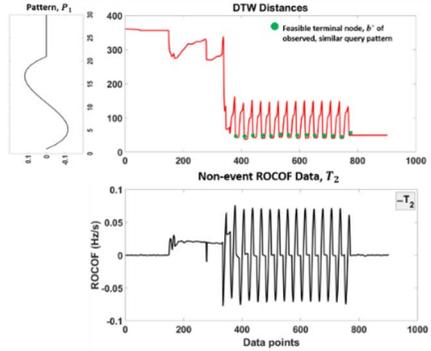


Fig. 11. ( $P_1, T_2$ )  $DTW$  distances in non-event  $T_2$

Fig. 11 shows the  $DTW$  distance at all data points for non-event test data  $T_2$ . New data points ( $\sim 0.25$  Hz/s) in non-event  $T_2$  are due to the window smoothing effect of the averaging event detection method used. The corresponding range of high  $DTW$  distance (280-330) is indicative of the absence of  $P_1$  data pattern in the initial subsequences of the test measurement. Further execution of the SDTW algorithm, and non-linear wrapping of  $P_1$  onto  $T_2$  leads to the identification of low  $DTW$  distance values ( $\sim 42$ ). Similar to Fig. 9, the detected subsequences of  $T_2$  are indicative of the presence of time errors. Here, the formulated accumulated cost matrix,  $AP$  is a much

wider matrix array ( $AP \in R^{30 \times 900}$ ), and shows 15 different instances of low cost neighborhoods where the time errors are most significant.

### c) Time-Errors with Noisy Signal

Real-life PMU data measurements are associated with noisy signals, thus we investigated the effect of noise on the pattern query search. A Gaussian-distributed noisy signal with signal-noise ratio (SNR) of 45-dB [26] was added to test ROCOF measurements during which the time error was observed for 5 seconds. The total simulation time was 10 seconds. Fig. 12 shows the  $DTW$  distance computed at every data point in  $T_3$ .

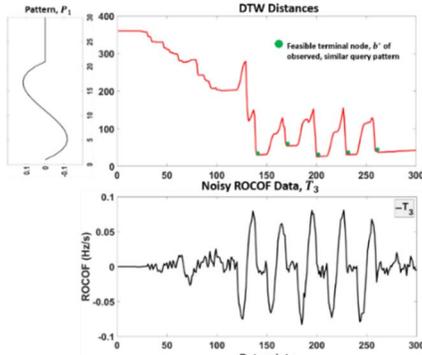


Fig. 12. ( $P_1, T_3$ )  $DTW$  distances in noisy data  $T_3$

A sudden roll-off of  $DTW$  distance values is observed for the early subsequences in  $T_3$ . However, the large magnitude of these computed values, and indicative of their non-correlation with the prototype data, eliminates the subsequences from being considered as instances of time errors. Further in the SDTW algorithm process, more significantly low  $DTW$  distances are observed. The identified subsequence patterns in the data string  $T_3$  possess different  $DTW$  distance values, that is,  $\sim (30, 55, 25, 31$  and  $41)$  though a similar phasor angle error is common to all the pattern instances found. The effect of noisy signal in the data stream is to introduce random increment (or reduction) in nodal costs associated with different warping paths. The result is a variation of  $DTW$  distance values for the same exact pattern observed at the different data string segments.

## V. CONCLUSION

In this paper, PMU time-related error detection using a dynamic time warping (DTW) variant technique on phasor angle second derivative measurements is proposed as an alternative to the sole dependence on PMU control information when quality issues are associated with flagged bits.

Different time mis-synchronization problems in PMU operation have been discussed. The analysis carried out for a clock delay error shows the use of a pre-defined, prototype ROCOF pattern to search for instances of time error in different test measurements supported by the computed similarity values. Since DTW algorithms are known to have quadratic time and space complexities, the proposed technique is suggested for data post-processing. Multiscale and approximation methods are proposed in future work to improve computation runtime.

## ACKNOWLEDGEMENTS

This research was supported by the Power System Engineering Research Center (PSERC) under project T-57.

## REFERENCES

- [1] D. Novosel, V. Madani, B. Bhargave, K. Vu, and J. Cole, "Dawn of the grid synchronization," *IEEE Power and Energy Magazine*, vol. 6, no. 1, pp. 49-60, 2008.
- [2] S. Chakrabarti, E. Kyriakides, T. Bi, D. Cai, and V. Terzija, "Measurements get together," *IEEE Power and Energy Magazine*, vol. 7, no. 1, pp. 41-49, 2009.
- [3] A. G. Phadke, "The Wide World of Wide-area Measurement," *IEEE Power and Energy Magazine*, vol. 6, no. 5, pp. 52-65, 2008.
- [4] A. G. Phadke and J. S. Thorp, "HISTORY AND APPLICATIONS OF PHASOR MEASUREMENTS," in *2006 IEEE PES Power Systems Conference and Exposition*, 2006, pp. 331-335.
- [5] "IEEE Standard for Synchrophasor Measurements for Power Systems," *IEEE Std C37.118.1-2011 (Revision of IEEE Std C37.118-2005)*, pp. 1-61, 2011.
- [6] Q. Zhang, L. Xiaochuan, D. Bertagnolli, S. Maslennikov, and B. Nubile, "PMU data validation at ISO New England," in *2013 IEEE Power & Energy Society General Meeting*, 2013, pp. 1-5.
- [7] T. Bi, J. Guo, K. Xu, L. Zhang, and Q. Yang, "The Impact of Time Synchronization Deviation on the Performance of Synchrophasor Measurements and Wide Area Damping Control," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1545-1552, 2017.
- [8] Q. Zhang, V. Vittal, G. Heydt, Y. Chakhchoukh, N. Logic, and S. Sturgill, "The time skew problem in PMU measurements," in *2012 IEEE Power and Energy Society General Meeting*, 2012, pp. 1-6.
- [9] Q. F. Zhang and V. M. Venkatasubramanian, "Synchrophasor time skew: Formulation, detection and correction," in *2014 North American Power Symposium (NAPS)*, 2014, pp. 1-6.
- [10] W. Yao *et al.*, "Impact of GPS signal loss and its mitigation in power system synchronized measurement devices," in *2017 IEEE Power & Energy Society General Meeting*, 2017, pp. 1-1.
- [11] C. Huang *et al.*, "Data quality issues for synchrophasor applications Part I: a review," *Journal of Modern Power Systems and Clean Energy*, vol. 4, no. 3, pp. 342-352, 2016.
- [12] D. P. Shepard, T. E. Humphreys, and A. A. Fansler, "Evaluation of the vulnerability of phasor measurement units to GPS spoofing attacks," *International Journal of Critical Infrastructure Protection*, vol. 5, no. 3-4, pp. 146-153, 2012.
- [13] A. Bose, "Smart Transmission Grid Applications and Their Supporting Infrastructure," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 11-19, 2010.
- [14] "IEEE Standard for Synchrophasor Data Transfer for Power Systems," *IEEE Std C37.118.2-2011 (Revision of IEEE Std C37.118-2005)*, pp. 1-53, 2011.
- [15] P. NASPI, "PMU Data Quality: A Framework for the Attributes of PMU Data Quality and Quality Impacts to Synchrophasor Applications," 2017.
- [16] A. R. Goldstein, D. Anand, and Y. Li-Baboud, "Investigation of PMU Response to Leap Second: 2015," 2015.
- [17] J. Guo, "Data analytics and application developments based on synchrophasor measurements," 2016.
- [18] I. Idehen, Z. Mao, and T. Overbye, "An emulation environment for prototyping PMU data errors," in *2016 North American Power Symposium (NAPS)*, 2016, pp. 1-6.
- [19] T. Xu, A. B. Birchfield, K. S. Shetye, and T. J. Overbye, "Creation of synthetic electric grid models for transient stability studies," in *The 10th Bulk Power Systems Dynamics and Control Symposium (IREP 2017)*, 2017.
- [20] A. G. Phadke and J. S. Thorp, *Synchronized phasor measurements and their applications*. Springer, 2008.
- [21] C. M. Bishop, "Pattern recognition and machine learning (information science and statistics) springer-verlag new york," *Inc. Secaucus, NJ, USA*, 2006.
- [22] D. Rafiei and A. Mendelzon, "Similarity-based queries for time series data," in *ACM SIGMOD Record*, 1997, vol. 26, no. 2, pp. 13-25: ACM.
- [23] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and information systems*, vol. 7, no. 3, pp. 358-386, 2005.
- [24] M. Müller, *Information retrieval for music and motion*. Springer, 2007.
- [25] B. C. Giao and D. T. Anh, "Similarity search in multiple high speed time series streams under Dynamic Time Warping," in *Information and Computer Science (NICS), 2015 2nd National Foundation for Science and Technology Development Conference on*, 2015, pp. 82-87: IEEE.
- [26] M. Brown, M. Biswal, S. Brahma, S. J. Ranade, and H. Cao, "Characterizing and quantifying noise in PMU data," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1-5.