Systematic Study of Data Requirements and AMI Capabilities for Smart Meter Analytics

Kavya Ashok[1], Matthew J. Reno[2], Logan Blakely[2], Deepak Divan[1]
[1] Georgia Institute of Technology, Atlanta, Georgia 30332, USA
[2] Sandia National Laboratories, Albuquerque, New Mexico 87185, USA

***Abstract*— Timeseries power and voltage data recorded by electricity smart meters in the US have been shown to provide immense value to utilities when coupled with advanced analytics. However, Advanced Metering Infrastructure (AMI) has diverse characteristics depending on the utility implementing the meters. Currently, there are no specific guidelines for the parameters of data collection, such as measurement interval, that are considered optimal, and this continues to be an active area of research. This paper aims to review different grid edge, delay tolerant algorithms using AMI data and to identify the minimum granularity and type of data required to apply these algorithms to improve distribution system models. The primary focus of this report is on distribution system secondary circuit topology and parameter estimation (DSPE).**

***Index Terms*— AMI, correlation, linear regression, parameter estimation, noise, resolution, sampling interval, time synchronization**

I. Introduction

From 2009 to 2015, public and private investments under the US DOE-led smart grid initiatives totaled $7.9 billion, of which $5.2 billion have gone towards AMI (Advanced Metering Infrastructure) and customer metering technologies [1]. The primary objective driving AMI investments has been automated, timely, and accurate billing, which were previously hampered by weather and property access limitations. As of 2016, nearly half of the 152 million electricity customers in the US have smart meters and utilities have reported an average of $6 million savings per project per year from just remote billing and metering services. Additional smart meter functions include remote connect/disconnect, tamper detection, outage and voltage monitoring etc. which have had substantial grid benefits like providing customers with more control over their electricity consumption and lowering outage costs and customer inconveniences through faster outage restoration.

Thus, many utilities are committed to investing in this technology and the total number of smart meters is growing annually. However, despite their advantages, the current implementation of AMI infrastructure incur high costs; upwards of $300 per end-point [2] in addition to installation and commissioning. This has made AMI a highly expensive automatic meter reading (AMR) implementation and extracting maximum value from AMI data is critical.

Research over the past decade has shown that significant new value streams and applications, beyond just consumer billing, can be developed by coupling advanced analytics and AMI data. Time-stamped, granular voltage and power data from AMI has enabled numerous operational and planning capabilities such as load forecasting, topology reconstruction, phase identification, load dissagregation, etc. [7] - [20]. Many utilities have shown immense interest in integrating these functionalities within their O&M (Operations and Maintenance) systems or have already conducted pilot programs to test several of these algorithms.

However, AMI data has diverse characteristics depending on the utility implementing the meters. A lack of common metering practices or consistent data formats inhibits utilities from deriving maximum value from their AMI investments. This project aims to conduct a systematic study to identify the minimum data and AMI capabilities required to implement different algorithms and use-cases. This is an active area of research to develop guidelines for utilities who are looking to install smart meters or are trying to extract additional value from their existing AMI infrastructure.

Some of the key questions explored in this paper are:

- Sampling Time Interval: AMI meters record data at hourly intervals or less and make that data available to utilities or customers at least once a day [1]. While many utilities have opted for smaller time intervals such as 15 minutes, the additional benefit this unlocks, at the cost of higher bandwidth and communication requirements, is not completely evident.
- Average vs Instantaneous data: A programmable feature in smart meters is the ability to chose between average or instantaneous voltage readings. While the industry leans towards the former for billing purposes, there is still some ambiguity regarding this, and many utilities have expressed the need to understand which data type might be more beneficial.
- Reactive power information: Most smart meters today record voltage and active power data [5]. While state-of-the-art meters are equipped with advanced sensing capabilities and can provide additional information such as harmonics, load power factor, etc., many utilities have not enabled these features due to cost or other concerns.
- Data Resolution: In order to reduce data storage requirements, utilities may save compressed data with fewer decimal values from their voltage and power readings in their long-term databases. The impact of this loss of information and the additional benefit of storing higher resolution data is explored in this paper.

- Time Synchronization: Many smart meter algorithms depend on analyzing data aggregated from all smart meters on a feeder. Understanding the impact of time drifts within each meter's internal clock is an important aspect to consider.
- Non-controllable noise parameters: The authors have also studied the impact of non-controllable, but common sources of error, such as meter bias, random noise, and missing data on the AMI algorithms.

Section II explains the approach taken to test these algorithms and generate the synthetic AMI data set. Section III describes the different smart meter algorithms considered in this paper. Section IV discusses the results of the study on various noise parameters. Recommendations, future work and conclusions are included in Section V .

II. Approach and Simulation Setup

In order to determine what AMI capabilities and data features are required to enable different algorithms, the following steps were taken:

1 The most common state-of-the-art approach for each application, as seen in literature, was determined and replicated.
2 Synthetic AMI data was generated to test these algorithms by developing a realistic distribution feeder simulation in OpenDSS. This approach attempts to mimic real life implementation and performance of AMI algorithms.
3 A mathematical analysis to explain the experimental observations and understand the algorithms performance in the presence of various noise inputs was conducted.

The OpenDSS simulation is based on EPRI's Test circuit 'ckt5' [4]; a 12.47 kV residential feeder with 1379 residential loads and 584 transformers. 12 months of consumption data at 1-minute intervals was extracted from Pecan Street's [4] Dataport for the power flow simulation, and high-resolution voltage data was recorded at each consumer node. This consumption data was collected from homes located in Austin, Texas between 2013-2017. The synthetic data was tested on algorithms with different levels of noise and inaccuracies and their average accuracy reported for the entire feeder. Descriptions of the simulated feeder and load profiles are shown below in Fig. 1 to 3.

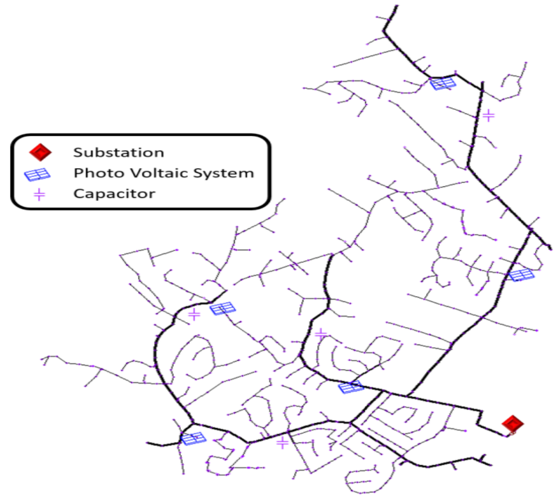

Fig. 1. Circuit plot of EPRI Ckt5

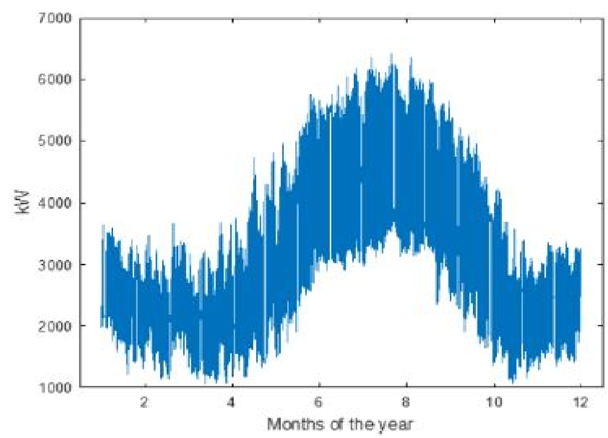

Fig. 2. Total feeder load during the yearlong simulation

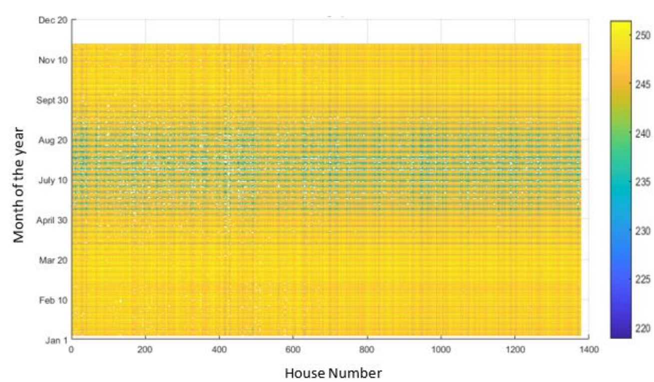

Fig. 3. Temporal raster plot of voltage (V) at each residential loads for a year at 15-minute resolution

III. Smart Meter Algorithms

A literature review of smart meter analytics was conducted to understand different use-cases for AMI data. The various applications, algorithm types and data requirements (in addition to GIS data) are summarized in Table I below. These are a mix of delay tolerant and real-time use-cases that can be enabled or improved using AMI data. Delay tolerant refers to those algorithms that can tolerate high latency in data retrieval and computation.

In this paper, the results have been explained with respect to topology correction algorithms, namely secondary network parameter estimation [7] and meter-transformer mapping [9]. The results for phase detection have been illustrated in [10]. The analysis for other algorithm categories such as rooftop solar, electric vehicle (EV) detection, non-technical loss detection, etc. is ongoing.

The exponential injection of new technologies such as electric vehicles, rooftop solar, storage, and other distributed energy resources (DERs) is changing the traditional behavior

TABLE I
SUMMARY OF SMART METER ANALYTICS

Category	Application	Data Requirements	Latency	References
Topology Correction/ Modeling	Phase Identification	Voltage magnitude	Delay tolerant	[12]
	Meter-Transformer Mapping	Voltage magnitude		[9]
	Secondary Network Parameter Estimation	Voltage magnitude, active & reactive power		[7]
Load Analysis	Load Modeling & Consumer Segmentation	Active Power, Weather data, Price Information		[19]
	Load Forecasting			[18]
	Load Disaggregation - Electric Vehicle Detection	Active Power	Delay tolerant/ Real-time	[17]
	Rooftop Solar Identification	Active Power, Voltage mag, Solar Irradiance		[20]
Power Loss	Detection of Non- technical losses	Load data, Voltage mag	Delay tolerant	[13]
	Outage Management	GIS, Voltage mag, Time of outage	Real time	[14]
	Fault Detection	Voltage magnitude, active & reactive power		[15]
Feeder Analysis	Detection of Low/High Voltage Areas	Voltage magnitude	Delay tolerant/ Real-time	[16]

and control of power systems. This increasing complexity in the distribution landscape requires utilities to enhance their visibility and to model their network with more granularity. Often, these connections change due to storms, zoning issues, etc. and might not get reflected in the utility's database. Incorrect or missing connectivity information in distribution system models of the low-voltage system can make it difficult for utility operators to take optimal operation decisions regarding the distribution system. Several works [13]-[20] have demonstrated that smart meter data can be utilized to correct these topological errors as explained below.

A. Meter-Transformer Mapping

The first step in developing the topology of low voltage networks is to determine which meters are connected to which phase or service transformer. Several papers have shown that the time-series voltage correlation between closely connected buses is much higher than with buses that are electrically further away [9]. While various approaches have been adopted, correlation and mutual information form the foundation of many of these algorithms. These have been shown to be effective with raw voltage data as well as a voltage fluctuation representation of the timeseries voltage data.

Fig. 4 illustrates this concept and shows voltage profiles of neighboring and non-neighboring buses.

The equation for Pearson correlation used on time series voltage data is shown in (1). Meters with a correlation coefficient higher than 0.97 are considered to have a nearby point of common coupling; likely to be the service transformer. Location information was utilized to ignore correlations between meters that were not within a 100m radius of each other.

$$R = \frac{\sum_{i=1}^{n}(V_{1i}-\overline{V_1})(V_{2i}-\overline{V_2})}{\sqrt{\sum_{i=1}^{n}(V_{1i}-\overline{V_1})^2 V_{2i}-\overline{V_2})^2}} \qquad (1)$$

where V_i refers to individual voltage readings for meters 1 and 2 and $\overline{V}$ is their respective voltage means.

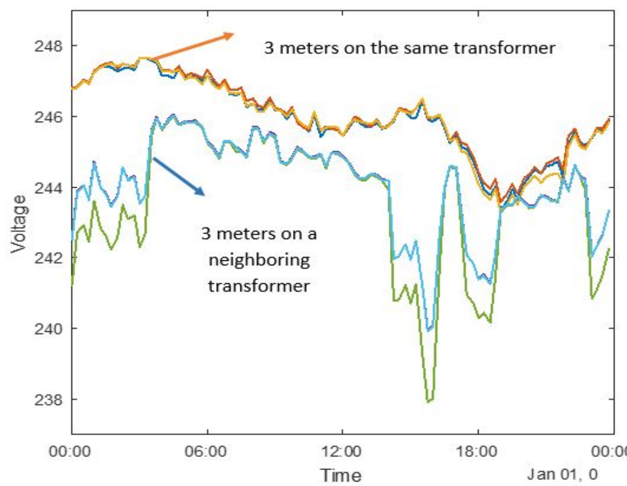

Fig. 4. Voltage profiles of neighboring and non-neighboring meters

B. Distribution System Secondary System Parameter Estimation

After pairing meters with their respective service transformer and phase, the exact topology and cable parameters of the low-voltage secondary network can be determined. The parameter estimation algorithm introduced in [12] and further developed in [7] and [8] distinguishes between serial and parallel connections and determines the length and effective impedance of cables from the service transformer to the load.

These algorithms employ linear regression and are based on the voltage drop approximation equation.

For Fig. 5, the equations for each parallel branch are:

$$V_o = V_1 + I_{R_1}R_1 + I_{X_1}X_1 + \epsilon_1 \qquad (2)$$

$$V_o = V_2 + I_{R_2}R_2 + I_{X_2}X_2 + \epsilon_2 \qquad (3)$$

$$\ldots$$

$$V_o = V_n + I_{R_n}R_n + I_{X_n}X_n + \epsilon_n \qquad (4)$$

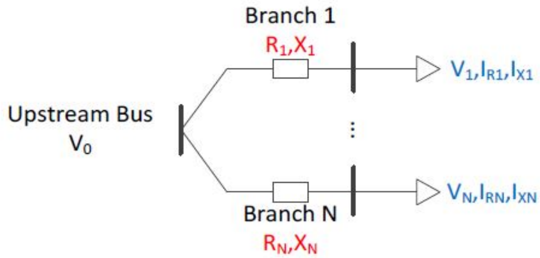

Fig. 5. Radial distribution system secondary system parameter estimation. The measured values (blue) are used to estimate the circuit topology and impedance (red) [7]

where $I_R = P/V$, $I_X = Q/V$;

P and Q are load active and reactive power. This can be rewritten in the form of a linear regression problem:

$$V_1 - V_2 = V = I_{R_2}R_2 + I_{X_2}X_2 - I_{R_1}R_1 - I_{X_1}X_1 \quad (5)$$

The results of both the algorithms, applied on the entire circuit consisting of 1379 residential customers, are summarized in Table II. Single phase meters were correctly paired with other meters on the same service transformer with an accuracy of 93.8%. Cable resistance (R) and reactance (X) were accurately estimated with a mean error of 1.2% and 2.6% respectively. If the accurate cable type is known, this equates to an average error of 1.26 ft. These results have been used as reference for subsequent analysis.

TABLE II
SUMMARY OF ALGORITHM ERRORS FOR CKT5 TEST SYSTEM

Meter-Transformer Mapping			
Accuracy	93.8%		
Parameter Estimation			
	R% error	X% error	Length (ft)
Mean Error	1.2064 %	2.6187%	1.26
Mean Absolute Error	1.5400 %	3.284%	1.4
Root Mean Square Error	1.653 %	3.5010%	1.5

IV. RESULTS AND DISCUSSION

The experiment design and results for different cases have been summarized below. Corresponding trends can be observed in Fig.6 for error in pairing transformers and Mean Absolute Error of all resistance estimations.

1) **Sampling Time Interval**: The 1-minute granularity time series measurements are grouped by the new interval size and then averaged (or the last reading is taken for instantaneous voltage measurements) to obtain the new timeseries measurement for 5, 15, 30 and 60 minutes.
   Observations: Fig.6a shows that when the data is perfect, i.e. no noise is added, the time interval does not impact accuracy. Averaged voltage data performs better than its instantaneous counterparts for larger intervals.
2) **Data Resolution:** Each measurement for the required time interval data set is rounded to the desired resolution. The maximum resolution provided by the simulation is 10mV and 0.1W. This corresponds to 2 and 4 decimal points each.
   Observations: At least 0.1V and 0.1kW resolution (1 decimal point each) was required to maintain reasonable levels of accuracy as shown in Fig.6b. Finer resolution beyond that do not result in significant improvement.
3) **Time Synchronization:** A maximum offset from 0-4 minutes was added to each meter at the 1-minute interval resolution, and its impact was studied for the different time intervals.
   Observations: As the reporting time interval increases, the algorithm becomes more robust to time displacement errors (Fig.6c). This assumes the quantity of data is the same.
4) **Quantity of Data:** The impact of quantity of data was studied by utilizing data from different numbers of months.
   Observations: A small and steady increase in the accuracy was observed when more data is utilized. Fig.6d shows a 2-3% improvement in performance was observed between using 1 month and 1 year of perfect (noiseless) data.
5) **Meter Bias:** For each meter a bias error is chosen at random and added to each measurement of that meter. The bias ranged from 0-2% of the mean at intervals of 0.5% and sampled uniformly between the negative and positive of the maximum bias level.
   Observations: Fig.6e reveals that bias does not impact linear regression or correlation methods. For example, in the former it is absorbed as constant noise and only impacts the intercept and not the slope of the line.
6) **Uniform Noise:** Noise is injected into each meter reading at random within the range [-Max, +Max]. The maximum noise level is varied from 0 to 1% of the nominal at steps of 0.1% and the noise is distributed uniformly between the negative and positive of the maximum level.
   Observations: Noise noticeably affects the accuracy of the algorithm as seen in Fig.6f. Averaged and larger time intervals ($\geq$ 15 minutes) tend to be more robust to noise. 0.35-0.55% noise in voltage data is a reasonable threshold. For smaller time intervals ($\sim$ 5min) a significant increase in error is seen beyond the 0.35% threshold. For larger time intervals, while the increase in error is more linear the accuracy drops below 95% beyond the recommended noise range.
7) **Missing Data:** Random data points from the data sets were dropped after averaging to the required time intervals. This can be caused due to failure in the communication network, packet drops, outages etc.
   Observations: Missing data does not impact the algorithm performance in both cases if sufficient data is used (here, 4-6 months). This is possibly because

TABLE III

SUMMARY OF AMI DATA REQUIREMENTS

Parameter	Without Base Noise	With Base Noise
Measurement Interval	Any interval	30 and 60 min.
Average Vs Instantaneous	Any type for smaller intervals. Average for larger intervals (>30min)	Averaged data performs well especially in the presence of time synchronization issues.
Data Resolution	At least 0.1V and 0.1kW (1 decimal) required.	
Meter Bias	No Impact	
Meter Precision(Noise)	0.35-0.55 % maximum noise in Voltage : Corresponds to ANSI standards class 0.2	
Time Synchronization	Averaged and larger measurement intervals are more robust	
Missing Data	Low sensitivity to missing data	
Data Quantity	$\sim$ 1 month	Error decreases steadily with larger quantities of data.

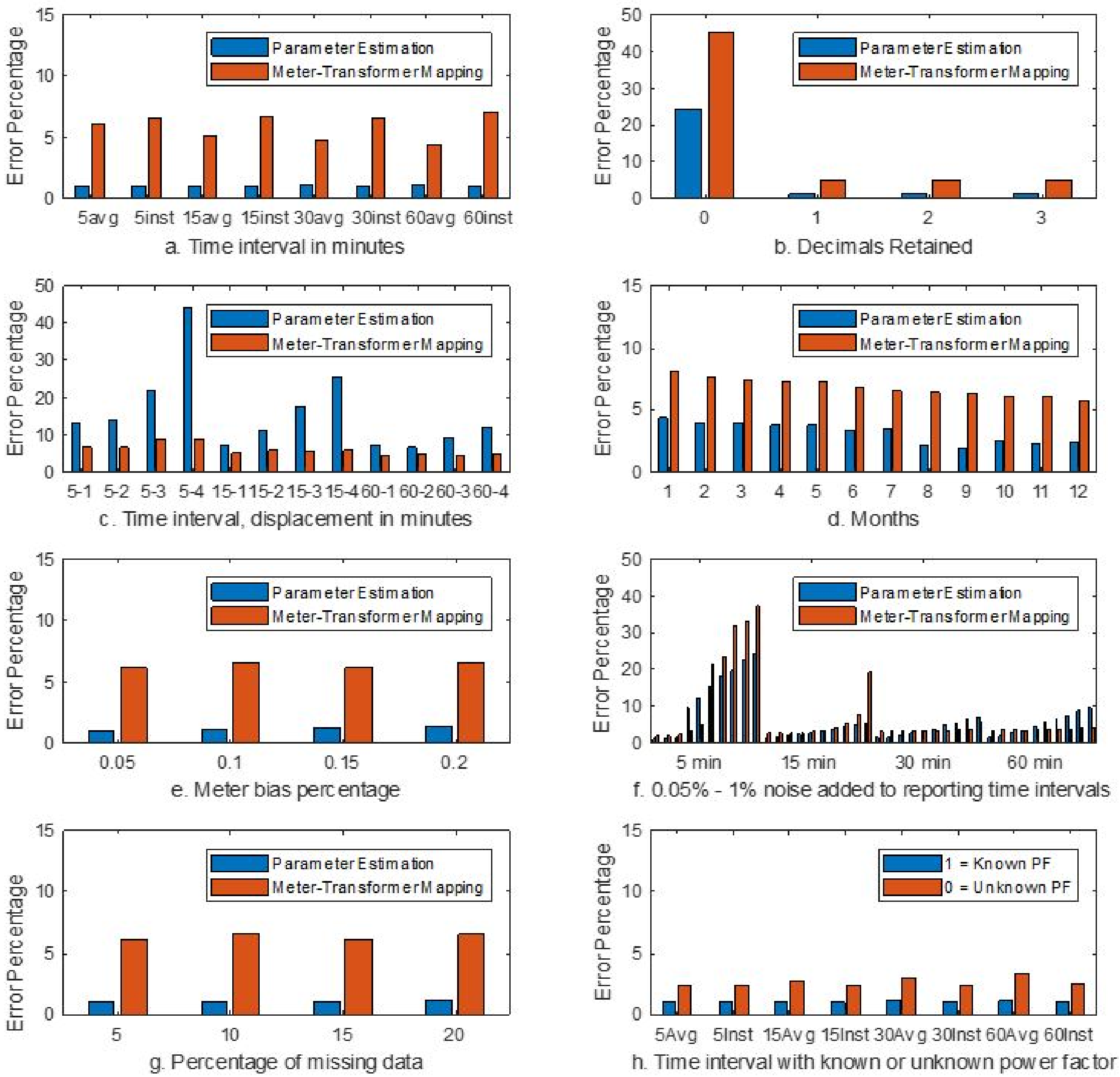

Fig. 6. Algorithm Accuracy Results when varying: a) Reporting Type and Measurement Interval, b) Data Resolution, c) Time Synchronization, d) Quantity of Data, e) Meter Bias, f) Uniform Noise, g) Missing Data, and h) Unknown Power factor

linear regression and correlation can perform well in the presence of missing data.

8) **Unknown Power Factor:** Often, smart meters do not record kVAr, and there is no power factor information available. Many algorithms based on AMI data, such as parameter estimation described in Section III, require kVAr readings. To get around this, random power factors were employed. Additionaly, the value of knowing the type of cable (here, X/R ratio= 0.304) was explored.

   Observations: Knowledge of the cable type used by the utility is extremely powerful. Fig.6h compares the results of DSPE when kVArh was measured (indicated by 1) vs when kVArh was not measured but the cable's X/R ratio was utilized (indicated by 0). Using a fixed X/R ratio, i.e. constant line slope in the linear regression formulation, the algorithm performs as well as the original case.

While Fig.6 compares the impacts of testing a range of individual error cases, these scenarios might not provide sufficient insight into optimal AMI capabilities as many of these errors occur simultaneously. Thus, to mimic realistic data with reasonable error levels, the following errors were injected into the 1-minute interval base data set. The noise levels conform with meter standards such as ANSI C12.

- Data Resolution – 0.1 V
- Meter Bias – 0.2%
- Meter Precision (Noise) – 0.2%
- Time Synchronization Issues — 1 min displacement for 20% of meters

The results from the simulations with multiple errors simultaneously injected show similar overall trends to the simulations of the errors in isolation from Fig. 6. Table III shows combined recommendations for topology estimation algorithms using smart meter data based on the results from testing the AMI data parameters and errors in isolation as well as comparing with the results from injecting errors simultaneously.

V. CONCLUSIONS

The type and granularity of AMI data collected are important considerations for implementing smart grid algorithms such as parameter estimation, phase detection etc. This paper reviews various use-cases and makes recommendations on appropriate data types and AMI capabilities. The additional value derived from different data types and resolutions has also been analyzed. The paper experiments with both controllable parameters such as measurement intervals, data resolution, etc., as well as uncontrollable features such as meter bias. It was concluded that averaged data at 30 or 60 minute intervals with 0.1V and 0.1kW resolution was optimal for topology related algorithms. While smaller intervals like 1 or 5 minutes provide more granularity and less data loss, they were also less robust to noise and other issues.

Algorithms not covered in this paper might have different data requirements. For example, load dissagregation might require data at higher sampling intervals, but time synchronization issues might not affect its accuracy. Thus, more use-cases need to be tested with a similar approach in the future. This study will allow utilities to determine optimal smart meter features that drive maximum value for them.

REFERENCES

[1] US Department of Energy, "AMI and Customer Systems: Results from the SGIG Program", 2016 [Online].

[2] "Advanced Metering Count by Technology Type" [Online]. Available: $https://www.eia.gov/electricity/annual/html/epa\_10\_10.html$

[3] "NEETRAC 16-091 'Assessment of Voltage Volality at the Grid Edge'-Final Report"

[4] EPRI Circuits [Online]. Available: http://svn.code.sf.net/p/electricdss/code/trunk/Distrib/EPRITestCircuits

[5] Itron Smart Meter.[Online] Available :'https://www1.itron.com/local/Sweden%20Product%20Portfolio/EM420i%20EN%2002-14.pdf'

[6] ANSI 12.20-2010 for Electricity Meters - 0.2 and 0.5 Accuracy Classes. American National Standard Institute (ANSI)

[7] J. Peppanen, M. J. Reno, R. J. Broderick, and S. Grijalva, "Distribution System Secondary Circuit Parameter Estimation for Model Calibration," Sandia National Laboratories, SAND2015-7477, 2015.

[8] M. Lave, M. J. Reno, R. J. Broderick, and J. Peppanen, "Full-Scale Demonstration of Distribution System Parameter Estimation to Improve Low-Voltage Circuit Models," IEEE Photovoltaic Specialists Conference (PVSC), 2017.

[9] K. Ashok, D. Deepak, and F. Lambert. "Grid edge analytics platform with AMI data." 2018 IEEE power & energy society innovative smart grid technologies conference (ISGT). IEEE, 2018.

[10] L. Blakely, M. J. Reno, K. Ashok, "AMI Data Quality and Collection Method Considerations for Improving the Accuracy of Distribution Models," IEEE Photovolt. Spec. Conf. (PVSC), 2019.

[11] L. Blakely, M. J. Reno, "Identifying Common Errors in Distribution System Models." IEEE Photovoltaic Specialists Conference (PVSC), Forthcoming 2019

[12] T. A. Short, "Advanced metering for phase identification transformer identification and secondary modeling", IEEE Transactions on Smart Grid, vol. 4, no. 2, pp. 651-658, 2013.

[13] P. Jokar, N. Arianpoo and V. C. M. Leung, "Electricity Theft Detection in AMI Using Customers Consumption Patterns," in IEEE Transactions on Smart Grid, vol. 7, no. 1, pp. 216-226, Jan. 2016.

[14] K. Kuroda, R. Yokoyama, D. Kobayashi, and T. Ichimura, "An approach to outage location prediction utilizing smart metering data," i8th Asia Model. Symp. (AMS), Taipei, Taiwan, 2014, pp. 6166.

[15] W. Luan, W. Xu, B. Harapnuk, "Method for identifying a system anomaly in a power distribution system", U.S. Patent 20150316620 A1, Nov., 2015.

[16] B. Wang, W. Xu, Z. Pan, "Voltage sag state estimation for power distribution systems," IEEE Trans. Power Systems, vol. 20, 2005.

[17] Z. Zhang, J. H. Son, Y. Li, M. Trayer, Z. Pi, D. T. Hwang, J. K. Moon, "Training-Free Non-Intrusive Load Monitoring of Electric Vehicle Charging with Low Sampling Rate," The 40th Annual Conference of the IEEE Industrial Electronics Society (IECON 2014), 2014.

[18] R. E. Edwards, J. New, and L. E. Parker, "Predicting future hourly residential electrical consumption: A machine learning case study," Energy Build., vol. 49, pp. 591603, Jun. 2012.

[19] J. Kwac, J. Flora, and R. Rajagopal. "Household energy consumption segmentation using hourly data." IEEE Transactions on Smart Grid, 2014.

[20] X. Zhang, and S. Grijalva. "A data-driven approach for detection and estimation of residential PV installations." IEEE Transactions on Smart Grid, 2016.

[21] L. Blakely, M. J. Reno, W. Feng, "Spectral Clustering for Customer Phase Identification Using AMI Voltage Timeseries", IEEE Power and Energy Conference at Illinois (PECI), 2019.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energys National Nuclear Security Administration under contract DE-NA0003525.