Applications of Big Data in Power Systems- A review

[1]Rohit Gupta, [2]Dr. KT Chaturvedi
[1]Phd Scholar, [2]Associate Professor
Department of Electrical Engineering
UIT RGPV Bhopal

ABSTRACT

Power system is a highly interconnected network which delivers electric power to electricity users. The size of this network varies from small isolated power systems to very large networks like the Power Grid. Sustaining the secure and reliable delivery of electric power requires continuous monitoring of the system. Measuring devices are installed at many locations in the system and the power characteristics including voltage, current, phase angle, and frequency are recorded continuously. To process the large volumes of data obtained from these measuring devices, it is essential to investigate effective data enhancement techniques. Various aspects of data mining including clustering, classification, regression and pattern finding are utilized to effectively manage the data and extract useful information from it.
Neural networks (NNET), support vector machine (SVM), k-means clustering, decision tree and visualization techniques are some of the popular algorithms in data mining which are frequently integrated with the power system studies. Application of an integrating technique combining two or more of these algorithms is also common in power system studies. Data mining offers effective solutions with low computation complexity and high performance to challenging problems in many field of power system including, but not limited to, stability analysis, fault detection, catastrophe prediction, load forecasting, and power system visualization.
Actually, Big Data technology has already been successfully applied as a powerful data-driven tool for solving numerous new challenges in power grid, such as price forecasting [7, 8], load forecasting [9], transient stability assessment [10], outlier detection [11], and fault detection and analysis [12], among others [13, 14]. We also want to get the rectified output from the Big data Technology and work on following areas:
Price Forecasting
Load Forecasting
Transient stability assessment
Outlier detection
Fault detection and analysis

Keywords: Big Data analytics,fault detection, high-dimensional data, stream data mining, nonlinear data

INTRODUCTION

In the course of the most recent decade, the world confronted an advanced transformation, and the expanded digitalization of various parts, together with the quick improvement of the Internet of Things (IoT) idea and an expansion in the quantity of associated gadgets brought about the aggregation of a lot of information. This is as yet developing and it is normal that by 2020 the information made and traded will achieve 44 ZB.

In the vitality segment, with the organization of the idea of keen frameworks and the expanding entrance of Information and Communication advances (ICT), digitalization is a reality. This digitalization can be found in all means of the vitality worth chain, from age to circulation/retail [2]. In parallel to it, control frameworks are confronting various moves identified with operational angles, for example, the requirements to decrease operational expenses and increment productivity, increment the sustainable power source share and think about natural issues, among other. These difficulties, together with the developments brought by enormous information speak to new chances, especially on the grounds that the up and coming age of intensity frameworks, the shrewd lattices, will be amazingly information escalated [3]–[6]. At last, the use of huge information methods in power frameworks can prompt continuous enhancement of intensity age and transmission, precise forecast of burden request, utilization designs examination prompting new administrations, and dynamic evaluating techniques [7]. Notwithstanding, current status of both power frameworks and huge information advancements still need to conquer various difficulties to meet these objectives. Most significant difficulties include: compelling information procurement and capacity; information curation strategies; how to precisely utilize the put away information to extricate business worth and how to moderate security while utilizing these information. These difficulties become significantly progressively significant in an area which is customarily unfriendly to change which is the situation of the vitality segment. Two of the primary explanations behind this conduct are dependability and long return-of-venture (ROI) situations. The vitality segment must guarantee a high unwavering quality and once an innovation is demonstrated to be protected from the operational perspective most utilities are hesitant to change to another innovation. Furthermore, this is an advantage concentrated part where most resources have a surprising expense and an enormous ROI which demonstrate that these equivalent resources must guarantee a protected task for long time. Thinking about this, the reception of new innovations and the change to an information driven methodology must be performed in an auspicious way and utilizing openings that emerge from the difficulties the area is confronting.

So as to more readily see how enormous information can be connected to control frameworks and how these can profit by methods which are as of now basic to the ICT world yet observed as abnormal to the power frameworks world, this paper introduces a short rundown of potential uses of huge information in power frameworks. To achieve an increasingly solid status, a reasonable model is demonstrated mulling over the use of an information driven way to deal with survey the sea vitality potential for wave vitality transformation. This methodology is by and large right now pursued under the system of a H2020 task named BigDataOcean. The paper is sorted out as pursues: area II depicts without further ado the primary qualities of huge information and how these can be adjusted to fit the requirements of intensity framework; segment III portrays how enormous information strategies can be utilized to play out an appraisal of sea vitality potential, centered in the wave vitality transformation frameworks; from that point forward, segment IV portrays the experiencing pilot of the undertaking BigDataOcean which is a concretization of the viewpoints referenced in the past two areas; at last, segment V contains a few remarks and ends identified with the paper subject.

BIG DATA

Big Data is also data but with a huge size. Big Data is a term used to describe a collection of data that is huge in size and yet growing exponentially with time. In short such data is so large and complex that none of the traditional data management tools are able to store it or process it efficiently.

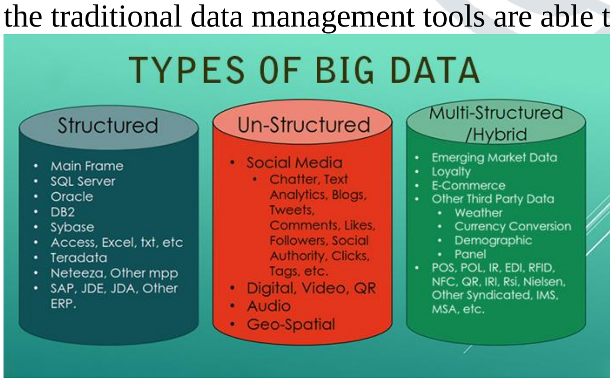


Fig. 1 Types Of Big Data

BRIEF REVIEW OF WORK ALREADY DONE IN THIS FIELD

Three main categories are identified for smart grid big data applications: Renewable Energy (RE), Demand Response (DR), and Electric Vehicles (EV) [2].

2.1. Renewable Energy

With expanding mix of sustainable power sources in power frameworks, information the board of current vitality lattices turns into an unpredictable errand, which ought to be tended to by huge information examination [28, 29]. For instance, verifiable climate information and GPS information can be utilized to improve anticipating of sustainable power source control age, which eventually upgrades the lattice vitality proficiency [30]. Information mining and preparing have been utilized to concentrate highlights of time arrangement information for increasingly exact guaging of irregular sustainable assets, for example, wind and sun oriented [31 - 34].

2.2. **Demand Response**
Request reaction alludes to changes in clients' power utilizations in light of changes in the power cost and accessibility [37]. Adaptable loads, for example, Heating, (HVAC), which "need to run yet their specific time of activity isn't basic" and other controllable loads, for example, electric vehicles are the objectives of interest reaction programs [38]. Conventional power frameworks don't offer continuous interest reaction, which debases lattice unwavering quality and ampleness. In this manner, huge information advances are utilized in savvy framework the executives to improve the power utilization information availability, which extends the interest reaction [39]. For instance, propelled meters apply game hypothesis and present day correspondence advances empowering shrewd matrices to give constant interest reaction capacity for increasingly productive and solid task of the matrix [40, 41].

2.3. **Electric Vehicles**
The International Energy Agency reports that more than 1.2 million Electric Vehicles (EVs) were working in 2015 [45] on the planet. In the US in 2015, 400,000 were working making around 1/3 of the world's absolute utilization of EV's.

EVs charge their batteries through the lattices, which forces a critical effect on electric matrix frameworks [46 - 48]. For instance, charging EVs in a populated zone during the pinnacle time may have outcomes, for example, intertwine victories, diminished productivity, and transformer debasement [49 - 51]. Through its bidirectional correspondence innovation, brilliant lattices can address these issues by booking the EV charging for off-top hours [52]. Likewise, by composed releasing through their vehicle-to-framework (V2G) capacities, EVs can give a few advantages, for example, auxiliary administrations, alleviating vulnerabilities of discontinuous sustainable power sources, for example, wind and sun based, and so on [53], [54 - 56].

There are a few investigations for organizing the EV charging/releasing to profit electric utilities and their clients utilizing hereditary calculations. EV driving and charging information have been broadly investigated by analysts to address the issues related with high entrances of EVs in electric networks. A group of specialists utilized an Estimation of Distribution Algorithms (EDAs) and populace based probabilistic hunt calculations to ideally deal with the colossal number of EV's charging [57]. Such calculations require the capacity to process immense and huge volume of constant information, which vigorously relies upon server-based handling or appropriated preparing systems. Another examination exhibited a system for EVs charging request utilizing enormous information investigation on information created by savvy meters [58]. Enormous information demonstrating for EV battery was proposed in [59] to improve estimation of driving extents with huge information distributed computing. Another examination exhibited basic leadership systems for EV charging by dissecting the anticipated age and request using line appropriations in a disseminated system [60]

3. **SMART GRID BIG DATA CHALLENGES AND PROPOSED SOLUTIONS**
Three main challenges are identified for big data in smart grids: security, quality, and processing location.

3.1. **Big Data Security**
The use of big data technology in smart grids substantially improves the network connectivity at the price of increased security vulnerabilities [61]. In a big data context, security exposures can be divided into three main parts: privacy, integrity, and authentication.

3.1.1. **Data Privacy**
Smart meters can be a main privacy concern if their data is not securely transferred and stored [62]. Smart meters gather control utilization information of network clients. Keen network suppliers break down such

information, which gives extraordinary instinct about clients' practices and propensities, to offer wise modified administrations [63]. A few techniques have been proposed to kill and limit the protection issue. These techniques incorporate, yet are not constrained to conveyed gradual information accumulation strategy [64], and veiling of utilization information implanted data [65]. Since the greater part of the current arrangements don't think about the tradeoff between expenses of lost security and information dispersal (utility), another technique is proposed to fulfill both protection and utility necessities of shrewd metered information [66].

3.1.2. Data Integrity

Risk of integrity attacks is a valid concern because any violation of integrity may cause security vulnerabilities [67]. Customer and network data are usually the targets for integrity attacks, and any modification of such data interrupts the data communication exchange and reduces the entire grid functionality [2]. For example, attackers can remove the higher degree nodes and replace them with higher probability nodes in the power network, which affects the integrity of data [67].

3.1.3. Data Authentication

Users in smart grids access the communication system through authentication, a process that verifies the user credentials against the accounts credential database [2]. Confirmation is utilized as an instrument to recognize substantial versus non-legitimate personalities inside most of existing security countermeasures [72]. One basic test that keen networks face is message infused assaults. On the off chance that such assaults are not tended to appropriately, they can altogether lessen the whole keen lattice execution [73]. To address such difficulties, a gathering of researchers proposed a validation strategy to verify brilliant lattice information correspondence trade with the utilization of Merkle hash-tree strategies [73]. Another examination proposed a protected message confirmation instrument by coordinating Diffie-Hellman conventions and hash-based message verification strategies [74]. Such structure permits shrewd meters inside the brilliant lattices to finish shared message confirmation assignments with insignificant sign trade and inertness [74].

3.2. Big Data Quality

Data quality refers to identifying and to removing the outliers before transferring the data to the system [75]. Energy power consumption data should have high degrees of quality to ensure correct data analysis and ultimately proper decisions. The quality issues of energy consumption data are categorized into noise data, incomplete data, and outlier data [76].

3.2.1. Noise Data

Generally, any data that is difficult to comprehend and/or to decode by computers is considered noise data, which degrades the data quality [76]. In a shrewd matrix setting, coherent mistakes and conflicting vitality utilization information are viewed as clamor [77, 78]. Coherent mistakes are characterized as the information that damages any given principles and attributes [79]. For instance, if the day by day client vitality utilization information incorporates 25 hours, it isn't intelligent as it surpasses the most extreme 24 hours [76]. Additionally, conflicting information happens when information does not pursue its recently concurred organization [80], or it needs sense when looking at its individual highlights [81, 82].

3.2.2. Incomplete Data

As the smart grid data complexity increases, incompleteness is occasionally observed in energy consumption data. Several methods such as delete tuple and data filing are developed to address incomplete data [82]. Delete tuple method simply removes the entire record with incomplete data. In any case, this strategy isn't fitting for situations where the informational collection has a few deficient perceptions [76]. In such cases, the fragmented information will be filled utilizing propelled calculations, for example, normal worth, counterfeit worth, and relapse examination [82].

3.2.3. Outlier Data

In statistics, if a point of data is considerably distant from other data points, it is called outlier [83]. In vitality utilization information, an anomaly might be treated as commotion and evacuated. Be that as it may, they may hold profitable data and along these lines, ought to be identified to protect the information quality. One strategy for identification is information quality mining, which is to review the information to consequently discover exceptions [84]. In savvy framework frameworks, anomalies ought to be recognized, distinguished, and dissected as they contain basic data, for example, control apportioning, gadget disappointments, and suspicious markers among others [85].

3.3. **Big Data Processing Location**

Preparing is a key capacity for using the calculations required by huge information. The present model for preparing is that data is collected and sent to a server farm to get handled and go to whomever needs the resultant data. The present structure as depicted by H. Jiang is the three-level structure with the fundamental information preparing at the inside with two layers around it for accumulation and appropriation [2]. There are go-between processors considered FOGs that are provincial accumulation focuses that likewise do negligible measures of preparing before passing its gathered data to the server farm [87].

Edge based preparing is turning into a bigger piece of the structure of huge information. With the drop-in cost to figure, scientists have begun to think back when processors had impediments and are making low power arrangements that can go anyplace and still have the option to process at any rate portions of an AI calculation on limited quantities of information. This makes the non-obtrusive burden estimating that is just made conceivable with low power installed frameworks [88].

PROPOSED METHODOLOGY

Although the energy domain data have been growing immensely, a majority of power system data is yet to be exploited. Many of energy domain legacy measurement devices and data management systems are based on the traditional concept of enterprise data warehousing, whereas under Big Data Analytic (BDA) approach some of its key fundamentals have been reconsidered. We identify several barriers and steps that need to be overcome/taken in power grids and microgrids to enable and facilitate BDA. We will also classify several core concepts, theories, and methods that could be leveraged in energy BDA. Finally, we will outline several of BDA research and application horizons in power system domain.

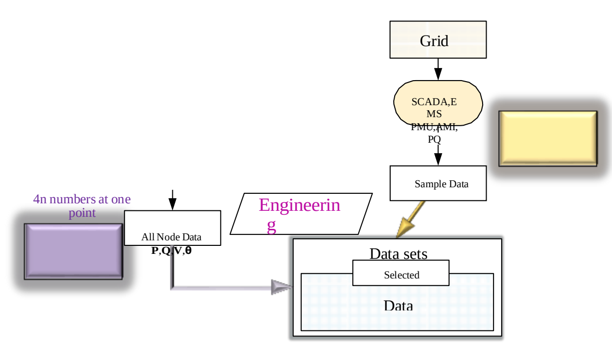
A. Big Data Architecture for Smart Grids

The frequently used notations are shown in Table.

Notations for architecture and case studies

Notations	*Means*
$\mathbf{A}_{\text{Time}}$	time area to focus on the data split-window
$\mathbf{A}_{\text{Node}}$	node area to focus on the data split-window
t	time: t_0 for current time, t_s for sampling time
$P_{\text{Bus-}n}$	power demand of load at bus-n
$P_{\max\ \text{Bus-}n}$	critical point on P_{ower}–V_{oltage} curve for bus-n
γ_{Acc}	addition factor for load fluctuations of the grid
γ_{Mul}	multiplication factor for load fluctuations of the grid

The designed architecture is illustrated as Fig. 2.



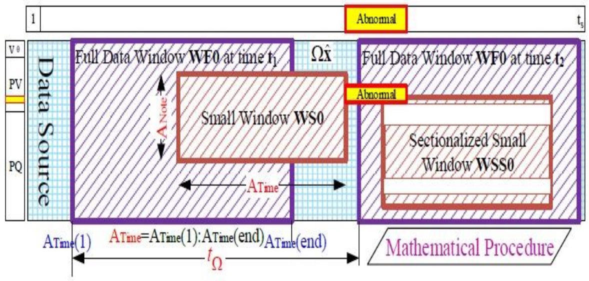


Fig. 2: The designed big data architecture for smart grids.

The part above the dot line illustrates an engineering procedure for big data modeling, during which the raw data source ^x are formed to map the physical system. The other part below the dot line illustrates a mathematical procedure for big data analysis. It is fully independent of engineering parameters, and during which the analyses are extracted from data source Ωxˆ .

It consists of two independent procedures to connect smart grids and big data—big data modeling as an engineering procedure, following by big data analysis as a mathematical procedure. During the engineering procedure, the raw data source Ωxˆ is acquired as described in Section II; during the mathematical procedure, following steps are conducted:

Steps of Mathematical Procedure

1) Set the initial parameters
   1a) Set $\mathbf{A}_{\mathrm{Time0}}$ and $\mathbf{A}_{\mathrm{Node0}}$ to focus on the first data window
   1b) Set t_Ω and $k=0$ to slide the moving split-window (MSW)
2) Focus on correspond window to form $\mathbf{X}$ ($\mathbf{A}_{\mathrm{Time}} = \mathbf{A}_{\mathrm{Time0}} + k$)
3) Calculate $\mathbf{X}$, $\mathbf{X}_u$, $\mathbf{Z}$, $\mathbf{Z}$, $\mathbf{S}$
4) Calculate the eigenvalues $\lambda_{\mathbf{Z}}, \lambda_{\mathbf{S}}$
5) Conduct ESD analysis and compare the result according to RMT
6) Calculate κ_{MSR} with $\lambda_{\mathbf{Z}}$
7) Visualize the results
8) Judge as times goes by:
   8a) $k < t_\Omega \Rightarrow k{+}{+}$; back to *step 2)*
   8b) k � $t_\Omega \Rightarrow$ END)

Especially, in step 2) Focus on the data window, we are able to conduct 1) real-time analysis: focusing on the real- time data window whose last edge of the sampling time area is current time (i.e. ATime(end)= t0); and 2) block-calculation for decoupling interconnected system: focusing on a smaller window consisting of data only in designated dimensions, but not in all dimensions. Besides, as the split-window, in a fixed size, slides across the data source Ωxˆ with tΩ and k set in 1b), a series of κMSR is got for further research and visualization.

B. Advantages in Data Processing

1) Algorithm: This architecture analyzes data in high-dimension as illustrated by the solid purple lines in Fig. 4. It is a universal procedure with 4 steps as follows:

Steps of Data Management for G3

1) Form standard random matrices $\tilde{\mathrm{X}}$
2) Acquire $\tilde{\mathrm{Z}}$, S by variables transforming ($\tilde{\mathrm{X}} \to \mathrm{X}_u \to \mathrm{Z} \to \tilde{\mathrm{Z}} \to \mathrm{S}$)
3) Conduct high-dimensional analysis based on RMT
4) Conduct engineering interpretations

On the other hand, the procedure of traditional data processing algorithms, in most cases, relies highly on specific simplifications and assumptions to build models. Taking genetic algorithm for an example, two steps are essential to achieve the result. One is to transcode the engineering variables to gene as the input of the gene model. It is a subjective selection procedure for the specific roles in engineering system, and only a few variables can be taken into account in final model. The other one is to perform the genetic algorithm through operations of selection, crossover, and mutation. Many problems, such as improper settings of the population size, of the crossover probabilities or of the mutation probabilities, will inevitably make the result worse.

1) Contrasted with customary calculations, huge information investigation, driven by information, empowers us to examine the interrelation and connection among every one of the components in framework, seen as relationship refered to by high-dimensional parameters. Utilizing an unadulterated metical strategy without physical models and hypo enormous information examination is simpler in rationale and quicker in Moreover, aside from stage 4): Conduct Engineering Interpret the entire technique is objective without presenting mulating the efficient blunders; besides, the inadvertent can be dispensed with. can be dispensed with either with the network size developing, repletion test and parallel figuring due to the indep of the calculation. *Distributed Calculation for Interconnected*

Smart grids operation are featured with autonomous and decentralized controls. Due to the potential scion cost and privacy concerns, Aggregating distribute sources to a centralized site for mining is system prohibitive. Then again, despite the fact that we can do show based mining at each appropriated site, the decoupling technique for associated sources is high identified with rearrangements and presumptions. In like manner the outcome are frequently scarcely fulfilled and prompts one-sided perspectives and choices. Enormous irregular lattices give a characteristic and widespread information driven arrangement. For a particular zone-separating interconnected framework, each site can shape a little grid just with its very own information. Along these lines, the incorporated framework can be isolated into squares for dispersed estimation and the other way around. For the general framework, we can direct high-dimensional investigation by coordinating the provincial lattices, or even by preparing a couple of local high-dimensional parameters. The scientific establishment is kept invariant as RMT; the adaptability, be that as it may, relies upon our goal. This engineering, decoupling the frameworks as measurable lattices or high-dimensional parameters rather than models, is viable for genuine enormous scale interconnected networks.

EXPECTED OUTCOME OF THE PROPOSED WORK

Many application areas of BDA in power systems are not unveiled yet. Nevertheless, several horizons of opportunities are expected already (Bui et al., 2012; Huang et al., 2014; Stimmel, 2014; Kezunovic et al., 2013; Moreno-Munoz et al., 2016; Hu and Vasilakos, 2016; Yu et al., 2015; Yin et al., 2013):

Enhanced Demand Response: Demand response, so far, has largely been the province of large utilities and large customers, due to the hindrance in management and value proposition for the operation of huge number of small loads. In any case, BDA will at long last empower utilities, or rather the cloud-based stages utilized by utilities, to take a gander at the utilization examples of a great many clients and quickly figure out which clients would partake in a DR occasion, how much these clients will charge for support, and how much will really be spared. A few approaches and devices are grown officially dependent on BDA that can fundamentally improve request reaction. For instance, Visualization and Insight System for Demand Operations and Management (VISDOM) is a stage for translation and extricating noteworthy data from enormous examples of keen meter information in a given utility administration territory or geographic locale (Kwac et al., 2014; Borgeson et al., 2015; Arguelles and Iglesias, 0000; Damström and Gerlitz, 2016). VISDOM is a collection of smart meter data analysis algorithms and visualization tools designed to address the challenge of interpreting patterns in energy data in support of utility energy efficiency and demand response programs. As one of it features, VISDOM allows to filter, sort, and assemble subsets of consumers using feature criteria and to display attributes of filtered customers using generic (i.e. histogram and scatter plot) or task specialized (i.e. cumulative sum and load shape) interactive visualizations (Borgeson et al., 2015).

Disaggregation and Fine Granularity Forecast: By utilizing BDA, estimates can be issued per-client for many clients like clockwork to tweak forecasts for power load over a whole district, in explicit geographic zones, or along specific appropriation feeder branches. The capacity to estimate each meter, transformer, and feeder will improve figure quality and spare in working expenses. For instance, anticipating apparatuses can be used in different applications, for example, power market costs, or client reaction (Kwac and Rajagopal, 2016; Balar et al., 2013; Gama and Rodrigues, 2007; Kezunovic et al., 2013). Clustering techniques have been proposed for determining natural segmentation of customers and identification of temporal consumption patterns (Kwac and Rajagopal, 2016; Balar et al., 2013). Price forecasting methods based on big price data have been proposed utilizing algorithms such as Grey Correlation Analysis, combination of Kernel function and Principle Component Analysis, and Support Vector Machine (Gama and Rodrigues, 2007). Architectures are also pro- posed, e.g. in Kezunovic et al. (2013), based on an online clustering algorithm and utilizing neural-network based predictive models.

Utilizing Domain and Off-Domain Data for Fault/Outage Detection: As knowledge is reached out down into circulation feeders, much information can be created and utilized for blackout identification and power rebuilding to enable utilities to enhance basic files. Shrewd meters record power utilization for charging, measure end-of-line voltage and, on account of a blackout, transmit a last heave as they lose control. With the across the board utilization of Social Media, comparative/complimentary data might be gotten from information sources, for example, clients' versatile tweets, that are connected to a location. For instance, examination, for example, neural systems, or fluffy rationales have been now used for ascertaining the flaw sift olds progressively so as to beat a portion of the difficulties in the customary security plans, especially for situations that issue is nonstationary, contains principal recurrence parts, or DC balances (Kezunovic et al., 2013; Vasilic and Kezunovic, 2005). As needs be, blackout location and power reclamation will be sped up.

Operations-Planning Convergence: This so-called convergence refers to the ability of a utility to realize the future conditions of power system with high probability and high accuracy. Operational planning refers to preparation for how weather, load, and generation conditions may change in the next minutes, hours, and days. This is difficult to achieve without systematic data management and accurate modeling. There are various reasons for this convergence gap, e.g. diverse models, diverse data sources and data for- mats, and inefficient data management tools, which all can be overcome with the unified methods and systematic data management. The challenges in this area are very broad. Some of the applications and approaches that are needed in this area include, but are not limited to: stochastic analysis, predictive analysis, dynamic scenario building for ‘‘what-if studies’’, and machine learning techniques to predict power system behaviors, capability to automatically recognize the changes in network topology due to planned and unplanned switching events, and ultimately enabling the power system to operate as interconnected subnetworks, that can be combined or separated, to achieve optimal power flows and improved reliability

Equipment Monitoring and Life Extension by Predictive Fault Detection: Prescient upkeep requires itemized data on the state of gear. Devoted condition checking frameworks, and additionally faculty to perform interim testing have been profoundly expensive for power framework gear that are differently situated on various pieces of the lattice. With BDA techniques notwithstanding, the nearby observing, corruption identification, and early disappointment anticipation on numerous framework resources can be accomplished. Also, the information from existing observing frameworks, that are conveyed for framework execution evaluation, can be utilized to furthermore serve for gear condition checking. For instance, the creators in Qiu et al. (2016, 2012), Feng et al. (2013), Qiu et al. (2017) and Long et al. (2015) create applications for encouraging precise breeze turbine disappointment recognition dependent on the turbine SCADA information. Note that, all huge utility scale turbines have a standard SCADA framework primarily utilized for execution checking. The creators additionally appeared in a SCADA sign contextual investigation on a 2 MW class variable-speed wind turbine that by observing gearbox oil temperature rise, control yield and rotational speed, a gearbox planetary stage disappointment could be anticipated and identified. Creating comparative applications for power framework hardware checking dependent on BDA could enormously diminish the O&M cost of the framework and the general ventures on new evaluate.

CONCLUSION

In this paper, we present an efficient and scalable platform aimed at helping domestic customers to save energy by managing their energy consumption positively. Improvements in the behavioral aspects in energy use have considerable potential. The users' own motivation also seems to play an
important role and thus, better results were achieved with customer involvement. The personal motivation for energy savings is based on different reasons and money saving is only one of them. Environmental concerns, governmental laws, social policies and technological restrictions are other powerful reasons where the future services should diversify in order to have greater impact. In addition, more encouraging and ad-hoc services must be provided to the customers.
Future work will analyze energy awareness depending on the place and nationality of the customers.

REFERENCES

[1] C.W. Gellings, M. Samotyj, and B. Howe, "The future's smart delivery system [electric power supply]", *IEEE Power Energy Mag.*, vol. 2, pp.40-48, 2004. [http://dx.doi.org/10.1109/MPAE.2004.1338121]

[2] H. Jiang, K. Wang, Y. Wang, M. GAO, and Y. Zhang , "Energy Big Data: A Survey,"", *IEEE Access*, vol. 4, pp. 3844-3861,2016. [http://dx.doi.org/10.1109/ACCESS.2016.2580581]

[3] J. Baek, Q. Vu, J. Liu, X. Huang, and Y. Xiang, "A secure cloud computing based framework for big data information management of smart grid", *IEEE Transaction on Cloud Computing*, vol. 3, no. 2, pp. 233-244, 2015. [http://dx.doi.org/10.1109/TCC.2014.2359460]

[4] B. Botte, V. Cannatelli, and S. Rogai, "The Telegestore project in ENEL's metering system", *Proc. 18th International Conference and Exhibition on Electricity Distribution*, pp. 1-4, 2005.

[5] Modern grid benefits, Conducted by National Energy Technology Laboratory for the U.S. Department of Energy, 2007.

[6] H. Farhangi, "The path of the smart grid", *IEEE Power Energy Mag.*, vol. 8, no. 1, pp. 18-28, 2010. [http://dx.doi.org/10.1109/MPE.2009.934876]

[7] Z. Bojkovic, and B. Bakmaz, "Smart grid communications architecture: A survey and challenges", *Proc. 11th International Conference on Applied Computers and Computational Science*, pp. 83-89, 2012.

[8] J. Duff, "Smart grid challenges", *Proc. Workshop on High Performance Transaction Systems (HPTS)*, 2009

[9] Z. Fan, P. Kulkarni, S. Gormus, C. Efthymiou, G. Kalogridis, and M. Sooriyabandara, "Smart grid communications: Overview of research challenges, solutions, and standardization activities", *IEEE Comm. Surv. and Tutor.*, vol. 15, no. 1, pp. 21-38, 2013. [http://dx.doi.org/10.1109/SURV.2011.122211.00021]

[10] How big data will make us more energy efficient, *Available* . https://www.weforum.org/agenda/2014/05/ big-data-will-make-us-energy-efficient/

[11] D.B. Rawat, and C. Bajracharya, "Cyber security for smart grid systems: Status, challenges and perspectives", *Proc. SoutheastCon 2015*, pp.1-6, 2015.

[12] J.N. Bharothu, M. Sridhar, and R.S. Rao, "A literature survey report on Smart Grid technologies", *Proc. 2014 International Conference on Smart Electric Grid (ISEG)*, pp. 1-8, 2014.

[13] N.B.M. Isa, T.C. Wei, and A.H.M. Yatim, "Smart grid technology: Communications, power electronics and control system", *Proc. 2015 International Conference on Sustainable Energy Engineering and Application (ICSEEA)*, pp. 10-14, 2015. [http://dx.doi.org/10.1109/ICSEEA.2015.7380737]

[14] C. Vineetha, and C. Babu, "Smart grid challenges, issues and solutions", *Proc. 2014 International Conference on Intelligent Green Building and Smart Grid (IGBSG)*, pp. 1-4, 2014.

[15] R. Mahmud, R. Vallakati, A. Mukherjee, P. Ranganathan, and A. Nejadpak, "A survey on smart grid metering infrastructures: Threats and solutions", *Proc. 2015 IEEE International Conference on Electro/Information Technology (EIT)*, pp. 386-391, 2015. [http://dx.doi.org/10.1109/EIT.2015.7293374]

[16] AT&T Smart Grid Technology Solutions, *Available*. https://www.business.att.com/enterprise/service/ internet-of-things/ smart-cities/iot- smart-grid/

[17] S. Sagiroglu, R. Terzi, Y. Canbay, and I. Colak, "Big data issues in smart grid system", *Proc. 5th International Conference on Renewable Energy Research and Applications*, pp. 1007-1012, 2016.

[18] M. Kezunovic, L. Xie, and S. Grijalva, The role of big data in improving power system operation and protection*Proc. Bulk Power System Dynamic and Control-IX Optimization, Security and Control of the Emerging Power Grid (IREP)*, 2013.

[19] M. Hardy, "Big data analytics - what it means to the audit community- part 1", *Available*.www.isaca.org/chapters2/jacksonville/events/Documents/ISACA%20Big%20Data%20for%20March%202013%E2%80%93%20Part%201.pdf

[20] T. Zhu, S. Xiao, Q. Zhang, Y. Gu, P. Yi, and Y. Li, "Emergent technologies in big data sensing: A survey", *Int. J. Distrib. Sens. Netw.*, vol.2015, pp. 1-13, 2015.

[21] C.K. Emani, N. Cullot, and C. Nicolle, "Understandable big data: A survey", *Comput. Sci. Rev.*, vol. 17, pp. 70-81, 2015. [http://dx.doi.org/10.1016/j.cosrev.2015.05.002]

[22] Big Data, Analytics, and Energy Consumption, *Available*. http://www.lavastorm.com/blog/post/big-data-analytics-and-energy-consumption/ [26] K.

Zhou, C. Fu, and S. Yang, "Big data driven smart energy management: From big data to big insights", *Renew. Sustain. Energy Rev.*, vol. 56, pp. 215-

225, 2016.[http://dx.doi.org/10.1016/j.rser.2015.11.050]

[23] Daniel Tunkelang talks about LinkedIn's data graph, *Available*.http://www.dataversity.net/daniel-tunkelang-talks-about-linkedins-data-graph/ [28] H. Al Haj Hassan, A. Pelov, and L. Nuaymi, "Integrating cellular networks, smart grid, and renewable energy: Analysis, architecture, and challenges", *IEEE Access*, vol. 3, pp. 2755-2770, 2015. [http://dx.doi.org/10.1109/ACCESS.2015.2507781]

[24] W. Tushar, J.A. Zhang, C. Yuen, D. Smith, and U. Naveed, "Hassan, "Management of Renewable Energy for A Shared Facility Controller in Smart Grid", *IEEE Access*, vol. 4, pp. 4269-4281, 2016. [http://dx.doi.org/10.1109/ACCESS.2016.2592509]

[25] J. M. Andujar, and F. Segura, "Study of a renewable energy sources-based smart grid. Requirements, targets and solutions, in Proc", In: *IEEE 3rd Conference on Power Engineering and Renewable Energy*,, 2016.

[26] M. Ghofrani, M. Ghayekhloo, and R. Azimi, "A novel soft computing framework for solar radiation forecasting", *Appl. Soft Comput.*, vol. 48, pp. 207-216, 2016. [http://dx.doi.org/10.1016/j.asoc.2016.07.022]

[27] R. Azimi, M. Ghofrani, and M. Ghayekhloo, "A hybrid wind power forecasting model based on data mining and wavelets analysis", *Energy Convers. Manage.*, vol. 127, pp. 208-225, 2016. [http://dx.doi.org/10.1016/j.enconman.2016.09.002]

[28] R. Azimi, M. Ghayekhloo, and M. Ghofrani, "A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting", *Energy Convers. Manage.*, vol. 118, pp. 331-344, 2016. [http://dx.doi.org/10.1016/j.enconman.2016.04.009]

[29] R. Billinton, and Y. Gao, "Multistate Wind Energy Conversion System Models for Adequacy Assessment of Generating Systems Incorporating Wind Energy", *IEEE Trans. Energ. Convers.*, vol. 23, no. 1, pp. 163-170, 2008. [http://dx.doi.org/10.1109/TEC.2006.882415]

[30] L. Kung, and H. Wang, "A recommender system for the optimal combination of energy resources with cost-benefit analysis", *Proc. International Conference on Industrial Engineering and Operations Management*, pp. 1-10, 2015.

[31] M.H. Albadi, and E.F. El-Saadany, "Demand response in electricity markets: An overview", *Proc. IEEE Power Engineering Society General Meeting*, 2007.

[32] A. Arabali, M. Ghofrani, M. Etezadi-Amoli, M.S. Fadali, and Y. Baghzouz, "Genetic-algorithm-based optimization approach for energy management", *IEEE Trans. Power Deliv.*, vol. 28, no. 1, pp. 162-170, 2012. [http://dx.doi.org/10.1109/TPWRD.2012.2219598]

[33] F. Rahimi, and A. Ipakchi, "Demand response as a market resource under the smart grid paradigm", *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 82-88, 2010. [http://dx.doi.org/10.1109/TSG.2010.2045906]

[34] S. Braithwait, and K. Eakin, "The role of demand response in electric power market design", *Edison Electric Institute, Available:*, 2002.http://www.eei.org

[35] Federal Energy Regulatory Commission, *FERC Order 719, Available:* http://www.ferc.gov/whats-new/comm meet/2008/ 101608/E-1.pdf [44] *American Recovery and Reinvestment Act of 2009, Available:*, 2009.https:// energy.gov/ sites/prod/ files/2014/ 12/f19/ SGIG-SGDP-Highlights-October2014.pdf

[36] International Energy Agency, "Global EV Outlook 2016: Beyond one million electric cars", *Available*.https:// www.iea.org/ publications/ freepublications/ publication/ Global_EV_Outlook_2016.pdf

[37] Z. Darabi, and M. Ferdowsi, "Aggregated impact of plug-in hybrid electric vehicles on electricity demand profile", *IEEE Transactions on Sustainable Energy*, vol. 2, no. 4, pp. 501-508, 2011. [http://dx.doi.org/10.1109/TSTE.2011.2158123]

[38] L.P. Fernandez, T.G.S. Roman, R. Cossent, C.M. Domingo, and P. Frias, "Assessment of the Impact of plug-in electric vehicles on distribution networks", *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 206-213, 2011. [http://dx.doi.org/10.1109/TPWRS.2010.2049133]

[39] Z. Jiang, L. Shalalfeh, and M.J. Beshir, "Impact of electric vehicle infrastructure on the city of chatsworth distribution system", *Proc. IEEE International Electric Vehicle Conference*, vol. vol. 1, pp. 17-19, 2014.

[40] R. Liu, L. Dow, and E. Liu, "A survey of PEV impacts on electric utilities", *Proc. IEEE PES Innovative Smart Grid Technologies*, pp. 1-8, 2011.

[41] E. Akhavan-Rezai, M.F. Shaaban, E. El-Saadany, and F. Karray, "Online intelligent demand management of plug-in electric vehicles in future smart parking lots", *IEEE Syst. J.*, vol. 10, no. 2, pp. 483-494, 2016. [http://dx.doi.org/10.1109/JSYST.2014.2349357]

[42] M.F. Shaaban, A.H. Osman, and M.S. Hassan, "Optimal coordination for electric vehicles in smart grids with high penetration of PV generation", *Proc. IEEE European Modelling Symposium*, 2016.

[43] S. Yoon, K. Park, and E. Hwang, "Connected electric vehicles for flexible vehicle-to-grid services", *Proc. IEEE International Conference on Information Networking*, pp. 411-413, 2017.

[44] E. Akhavan-Rezaei, M.F. Shaaban, E.F. El-Saadany, and F. Karray, "New EMS to Incorporate Smart Parking Lots into Demand Response",*IEEE Trans. Smart Grid*.

[45] M. Ghofrani, A. Arabali, and M. Ghayekhloo, "Optimal charging/discharging of grid-enabled electric vehicles for predictability enhancement of PV generation", *Electr. Power Syst. Res.*, vol. 117, pp. 134-142, 2014. [http://dx.doi.org/10.1016/j.epsr.2014.08.007]

[46] M. Ghofrani, A. Arabali, M. Etezadi-Amoli, and M.S. Fadali, "Smart scheduling and cost-benefit analysis of grid-enabled electric vehicles for wind

power integration", *IEEE Trans. Smart Grid*, vol. 5, no. 5, pp. 2306-2313, 2014. [http://dx.doi.org/10.1109/TSG.2014.2328976]

[47] W. Su, and M-Y. Chow, "Performance evaluation of an EDA-based large-scale plug-in hybrid electric vehicle charging algorithm", *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 308-315, 2012. [http://dx.doi.org/10.1109/TSG.2011.2151888]

[48] X. Zhang, and S. Grijalva, "An advanced data driven model for residential electric vehicle charging demand", *Proc. IEEE Power Energy Soc. General Meeting*, pp. 1-5, 2015.

[49] C.H Lee, and C.H. Wu, "A novel big data modeling method for improving driving range estimation of EVs", *IEEE Access*, vol. 3, pp. 1980-1993, 2015. [http://dx.doi.org/10.1109/ACCESS.2015.2492923]

[50] M.F. Shaaban, M. Ismail, E.F. El-Saadany, and W. Zhuang, "Real Time PEV charging/discharging coordination in smart distribution systems", *IEEE Trans. Smart Grid*, vol. 5, no. 4, pp. 1797-1807, 2014. [http://dx.doi.org/10.1109/TSG.2014.2311457]

[51] D. Dzung, M. Naedele, T.P.V. Hoff, and M. Crevatin, "Security for industrial communication systems", *Proc. IEEE*, vol. 93, no. 6, pp. 1152-1177, 2005. [http://dx.doi.org/10.1109/JPROC.2005.849714]

[52] C.S. Lai, and L.L. Lai, "Application of big data in smart grid", *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, pp. 665-670, 2015.

[53] P. McDaniel, and S. McLaughlin, "Security and privacy challenges in the smart grid", *IEEE Secur. Priv.*, vol. 7, no. 3, pp. 75-78, 2009. [http://dx.doi.org/10.1109/MSP.2009.76]

[54] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption", *Proc. 1st IEEE International Conference Smart Grid Communication*, pp. 327-332, 2010.

[55] G. Kalogridis, C. Efthymiou, S.Z. Denic, T.A. Lewis, and R. Cepeda, "Privacy for smart meters: Towards undetectable appliance load signatures", *Proc. 1st IEEE International Conference Smart Grid Communication*, pp. 232-237, 2010.

[56] S.R. Rajagopalan, L. Sankar, S. Mohajer, and H.V. Poor, "Smart meter privacy: A utility-privacy framework", *Proc. IEEE international Conference Smart Grid Communication*, pp. 190-195, 2011.

[57] S. Ruj, and A. Pal, "Analyzing cascading failures in smart grids under random and targeted attacks", *Proc. IEEE 28th International Conference Advanced Information Networking and Applications*, pp. 226-233, 2014.

[58] L. Xie, Y. Mo, and B. Sinopoli, "Integrity data attacks in power market operations", *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 659-666, 2011. [http://dx.doi.org/10.1109/TSG.2011.2161892]

[59] Y. Yuan, Z. Li, and K. Ren, "Quantitative analysis of load redistribution attacks in power systems", *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 9, pp. 1731-1738, 2012. [http://dx.doi.org/10.1109/TPDS.2012.58]

[60] O. Kosut, L. Jia, R.J. Thomas, and L. Tong, "Malicious data attacks on the smart grid", *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645-658, 2011. [http://dx.doi.org/10.1109/TSG.2011.2163807]

[61] L. Jia, R. Thomas, and L. Tong, "Malicious data attack on real-time electricity market", *Proc. IEEE International Conference on Acoustic, Speech, and Signal Processing*, pp. 5952-5955, 2011.

[62] H. Liu, H. Ning, Y. Zhang, and L.T. Yang, "Aggregated-proofs based privacy-preserving authentication for V2G networks in the smart grid", *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1722-1733, 2012. [http://dx.doi.org/10.1109/TSG.2012.2212730]

[63] H. Li, R. Lu, L. Zhou, B. Yang, and X. Shen, "An efficient Merkle-Tree-Based Authentication Scheme for Smart Grid", *IEEE Syst. J.*, vol. 8, no. 2, pp. 655-663, . [http://dx.doi.org/10.1109/JSYST.2013.2271537]

[64] M.M. Fouda, Z.M. Fadlullah, N. Kato, R. Lu, and X. Shen, "Towards a light-weight message authentication mechanism tailored for smart grid communications", *Proc. IEEE Conference on Computer Communications workshops*, pp. 1018-1023, 2011.

[65] B. Sasha, and D. Srivastava, "Data quality: The other face of Big Data", *Proc. IEEE 30th International Conference on Data Engineering*, 2014

[66] W. Chen, K. Zhou, S. Yang, and C. Wu, "Data quality of electricity consumption data in a smart grid environment", *Renew. Sustain. Energy Rev.*, vol. 75, pp. 98-105, 2017. [http://dx.doi.org/10.1016/j.rser.2016.10.054]

[67] S. Bai, X. Li, Z. Qin, and L. Zhu, "The application and research of Noise data acquisition with wireless network", *Proc. IEEE International Conference on Environmental Science and information*, vol. vol. 3, pp. 693-696, 2009.

[68] K. Hu, L. Li, and Z. Lu, "A cleaning method of noise data in RFID data streams", *Proc. 3rd International Conference on Consumer Electronics, Communications and Networks*, pp. 1-4, 2013.

[69] B. Fan, G. Zhang, and H. Li, "Multiple models fusion for pattern classification on noise data", *Proc. International Conference on System Science and Engineering*, pp. 64-68, 2012.

[70] E. Raham, and H.H. Do, "Data cleaning: Problems and current approaches", *Q. Bull. Comput. Soc. IEEE Tech. Comm. Data Eng.*, vol. 23, no. 4, pp. 3-13, 2000.

[71] A. Motro, P. Anokhin, and A.C. Acar, "Utility-based resolution of data inconsistencies", *Proc. ACM International Workshop on Information Quality in Information System*, vol. 76, pp. 35-43, 2004.

[72] J.R. Li, L.P. Khoo, and S.B. Tor, "RMINE: A rough set based data mining prototype for the reasoning of incomplete data in condition-based fault diagnosis", *J. Intell. Manuf.*, vol. 17, no. 1, pp. 163-176, 2006. [http://dx.doi.org/10.1007/s10845-005-5519-8]

[73] National Institute of Standards and Technology, *NIST/SEMATECH e-Handbook of Statistical Methods*.http:// www.itl.nist.gov/ div898/ handbook/

[74] D. Luebbers, U. Grimmer, and M. Jarke, "Systematic development of data mining-based data quality tools", *Proc. 29th International Conference on Very Large Data bases,* vol. 29, pp. 548-559, 2003.

[75] J. Zhang, and H. Wang, "A new pretreatment approach of eliminating abnormal data in discrete time series", *Proc. IEEE International Geoscience and Remote Sensing Symposium,* vol. 1, pp. 665-668, 2005.

[76] Y. Liu, P. Ning, and M.K. Reiter, "False data injection attacks against state estimation in electric power grids", *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, p. 13, 2009.

[77] F.Y. Okay, and S. Ozdemir, "A fog computing based smart grid model", *Proc. 2016 International Symposium on Networks, Computers and Communications (ISNCC),* pp. 1-6, 2016.

[78] T. Sirojan, T. Phung, and E. Ambikairajah, "Intelligent edge analytics for load identification in smart meters", *Proc. IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia),* pp. 1-5, 2017.

[79] B. Stojkoska, and K. Trivodaliev, "Enabling internet of things for smart homes through fog computing", *in 25th Telecommunication Forum (TELFOR),* pp. 1-4, 2017.

[80] Y. Zhang, K. Liang, S. Zhang, and Y. He, "Applications of edge computing in PIoT", *Proc. IEEE Conference on Energy Internet and Energy System Integration (EI2),* pp. 1-4, 2017.

[81] A. M. Annaswamy, M. Amin, A. M. Annaswamy, C. L. DeMarco, and T. Samad, IEEE Vision for Smart Grid Controls: 2030 and Beyond, Sept. 2013.

[82] DE-FOA-0001495, 2016. Enabling extreme real-time grid integration of solar energy (energise), Tech. rep., DEPARTMENT OF ENERGY (DOE).

[83] Sunshot vision study, February 2012. Tech. rep., Online access: http://energy.gov/ eere/sunshot/sunshot-vision-study.

[84] ICS-CERT, 2014. Ongoing sophisticated malware campaign compromising ics (up- date b), Tech. rep., Department of Homeland Security.

[85] Aiello, M., Pagani, G.A., 2014. The smart grid's data generating potentials, in: 2014 Federated Conference on Computer Science and Information Systems (FedCSIS), IEEE, pp. 9–16.

[86] Alejandro, L., Blair, C., Bloodgood, L., Khan, M., Lawless, M., Meehan, D., Schneider, P., Tsuji, K., 2014. Global market for smart electricity meters: Government policies driving strong growth. US International Trade Commission - Office of Industries. Andersen, M.P., Culler, D.E., 2016. Btrdb: Optimizing storage system design for timeseries processing, in: Proceedings of the 14th USENIX Conference on File and Storage Technologies (FAST 16).

[87] Andersen, M.P., Kumar, S., Brooks, C., von Meier, A., Culler, D.E., 2015. Distil: design and implementation of a scalable synchrophasor data processing system. In: 2015 IEEE International Conference on Smart Grid Communications (Smart- GridComm). IEEE, pp. 271–277.

[88] Ardakanian, O., Yuan, Y., Dobbe, R., von Meier, A., Low, S., Tomlin, C., 2017. Event detection and localization in distribution grids with phasor measurement units. arXiv preprint arXiv:1611.04653. Arguelles, D., Iglesias, R., Visual filtering of large energy consumption datasets by leveraging usage clusters.

[89] Arnold, D.B., Roberts, C., Ardakanian, O., Stewart, E.M., 2017. Synchrophasor data analytics in distribution grids. In: Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2017 IEEE. IEEE, pp. 1–5.

[90] Bai, Y., Zhong, H., Xia, Q., Kang, C., Xie, L., 2015. A decomposition method for network-constrained unit commitment with ac power flow constraints. Energy 88, 595–603.

[91] Balar, A., Malviya, N., Prasad, S., Gangurde, A., 2013. Forecasting consumer behavior with innovative value proposition for organizations using big data analytics. In: 2013 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). IEEE, pp. 1–4.

[92] Batra, N., Singh, A., Whitehouse, K., 2017. Neighbourhood nilm: A big-data approach to household energy disaggregation. arXiv preprint arXiv:1511.02900.

[93] Bernstein, A., Dall'Anese, E., 2017. Real-time feedback-based optimization of distri- bution grids: A unified approach. arXiv preprint arXiv:1711.01627.

[94] Bertsekas, D.P., Tsitsiklis, J.N., 1989. Parallel and Distributed Computation: Numer- ical Methods, Vol. 23. Prentice hall, Englewood Cliffs, NJ.

[95] Borgeson, S., Flora, J.A., Kwac, J., Tan, C.-W., Rajagopal, R., 2015. Learning from Hourly Household Energy Consumption: Extracting, Visualizing and Interpret- ing Household Smart Meter Data, in: International Conference of Design, User Experience, and Usability. Springer, pp. 337–345.

[96] Bui, N., Castellani, A.P., Casari, P., Zorzi, M., 2012. The internet of energy: A web- enabled smart grid system. IEEE Netw. 26 (4).

[97] Chen, H., Chiang, R.H., Storey, V.C., 2012. Business intelligence and analytics: From big data to big impact. MIS Q. 36 (4), 1165–1188.

[98] Chen, Y., Xie, L., Kumar, P.R., 2014. Power system event classification via dimen- sionality reduction of synchrophasor data. In: Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014 IEEE 8th. IEEE, pp. 57–60.

[99] Choi, J., Demmel, J., Dhillon, I., Dongarra, J., Ostrouchov, S., Petitet, A., Stanley, K., Walker, D., Whaley, R.C., 1996. Scalapack: A portable linear algebra library for distributed memory computers design issues and performance. Comput. Phys. Comm. 97 (12), 1–15.

[100] Chow, C.W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., Washom, B., 2011. Intra-hour forecasting with a total sky imager at the uc san diego solar energy testbed. Sol. Energy 85 (11), 2881–2893.