GRID MODERNIZATION INITIATIVE PEER REVIEW

GMLC 1.4.9 Integrated Multi Scale Data Analytics and Machine Learning

EMMA M STEWART – LAWRENCE LIVERMORE NATIONAL LABORATORY

MICHAEL CHERTKOV (+1) – LOS ALAMOS NATIONAL LABORATORY

April 18-20, 2017

Sheraton Pentagon City – Arlington, VA

Integrated Multi Scale Data Analytics and Machine Learning

High Level Summary

Project Description

Develop and demonstrate distributed analytics solutions to building to grid challenges, leveraging multi-scale data sets, from both sides of the meter.

Evaluate and demonstrate the application of machine learning techniques to create actionable information for grid and building operators, and derive customer benefits from disparate data

Project Objectives

- Enable local nodal information exchange and high-performance, distributed algorithmic analysis
- Deploy local analytics integration at the grid edge, building to grid interface, with a bridge to supervisory grid layers
- Leapfrog state-of-the-art strategies to accommodate DER and thrive in an evolving distribution system

Value Proposition

Cohesive view of the future distribution grid and its building interface, an interactive environment where there are consumer benefits and motivations to leverage customer behind-the-meter assets. Large spatial footprint of the distribution grid and diverse locations of its assets make **observability, monitoring and diagnosis of abnormal (faults) and even planned (demand response or DER dispatch) events** challenging tasks for the existing descriptive analytics field, but great for Machine Learning.

Integrated Multi Scale Data Analytics and Machine Learning

Project Team

GRID MODERNIZATION INITIATIVE
U.S. Department of Energy

LLNL: Lead Lab, LANL: +1

LLNL *– Data collection and application definition, ML for incipient failure and DR verification, distributed communications*

LBNL *– Platform development, incipient failure detection*

LANL – *Anomaly detection and platform integration*

ANL *– Distributed analytics for resiliency apps,*

NREL *– Application definition, ML for DER verification*

ORNL *– OpenFMB integration, platform review and selection, new sensor streams*

SNL *– Application development, topology detection*

PROJECT FUNDING			
Lab	FY16 $	FY17$	FY18 $
LBNL*	267	150	125
LANL	220	220	220
LLNL*	83	200	225
ANL	83	83	83
NREL	41.5	41.5	41.5
SNL	41.5	41.5	41.5
ORNL	104	104	104

*budgets being reorganized due to change in personnel

Industrial Partners*: PSL, National Instruments, OSISoft, SGS, Sentient, SGIP*
Utility Partners*: Riverside Public Utility, Pecan Street/Austin Energy, PG&E, Duke Energy*

Integrated Multi Scale Data Analytics and Machine Learning

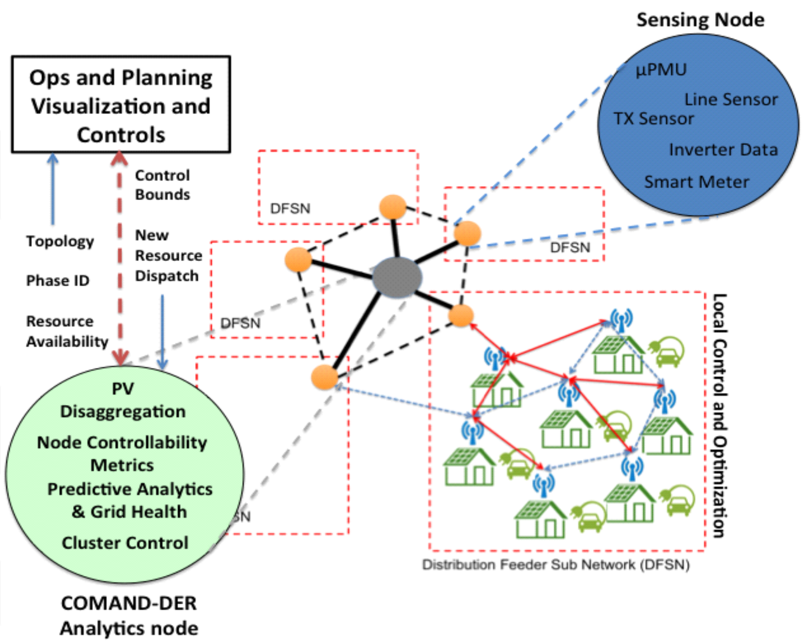
Relationship to Grid Modernization MYPP

- Devices and Integrated System Testing
  - MYPP Activity 3
    - Task 2.3.8 - Testing and validation for transactive energy
- Sensors and Measurement
  - MYPP Activity 1 - Improved Sensing for Buildings
    - Task 3.1.1 New low cost building sensors
  - MYPP Activity 2 - Enhanced sensing for distribution
    - Task 3.2.3 Connectivity/ Topology discovery
    - Task 3.2.4 & 3.2.5 New sensors for asset and state
  - **MYPP Activity 4 – Data Analytics and Visualization**
    - **Task 3.4.1 Develop real time streaming analytics and ML paradigms for grid visibility control resiliency and security**
    - Task 3.4.2 Distributed Analytics Tools
    - Task 3.4.3 Heterogeneous data collection for EGS
    - Task 3.4.4 Dynamic behavior prediction
- System Ops P Flow and Control
  - MYPP Activities 1,3,4
    - Task 4.1.2 & 3 Control Theory & Validation
    - Task 4.3.1 Uncertainty Apps
- Security and Resilience
  - MYPP Activity 2 & 3
    - Task 6.3.1 data ingest
    - Task 6.3.3 Cyber threat detection

Integrated Multi Scale Data Analytics and Machine Learning

Approach Overall

- 1) Setting the stage
- 2) Evaluation and testing demo of state of art in Distributed ML
- 3) Stakeholder demonstration at metrics evaluation
- 4) New ML technique development and application
- 5) Coordinated project integration



- In year 1 we are illustrating R & D analytics white space with application of both existing and new techniques
  - Benefits to consumers and utilities at the building to grid interface

Unique Aspects of approach:

- **Streaming data demonstrated in field,**
- **Distributed and in data in motion, as opposed to centralized**
- **Novel algorithms to be applied at building to grid interface**

Integrated Multi Scale Data Analytics and Machine Learning

Approach 1: White Paper & Use Case Development

GRID MODERNIZATION INITIATIVE
U.S. Department of Energy

- White Paper Goals
  - What is machine learning and why do we need it from two perspectives building/grid and data science
  - Illustrate the potential for application development
    - Where we can improve and innovate?
    - Where can we improve existing techniques with new data?
  - This will enable value streams to be derived from new sensing and grid architecture for many years to come
    - Outline a framework to define clear benefits to consumers and utilities

Integrated Multi Scale Data Analytics and Machine Learning

Approach 2 – Case Study Identification and Review

Use Case	DR & DER Local Availability & Verification	Incipient Failure Detection in Distribution	Topology & Parameter Estimation
Present State of Art	Estimated forecast and manual communication	Local sensing, smoke signals, outage management	Successful applications in highly sensed environments,
Present Granularity	Sub or Individual Customer, Day+	Limited prior to outaged component	Sub or Individual Customer, Day+
Future Requirement	Cust & Dist XFRMR Real time and Hrs Ahead	Dist XFRMR/ component level Real time, Months and Hrs Ahead	Switch, Distribution Component Planning and Event Driven
Useful Data	AMI, Irradiance, Green/Orange Button, PMU, model	AMI, Model, PMU, GIS	AMI, Model PMU, GIS, Model
Stakeholders	Consumers, DERMS and PV Vendors, Operators	Consumers, Asset Managers, Operators	Planners, vendors, PV integrators, Operators

Integrated Multi Scale Data Analytics and Machine Learning

Approach 3 – Application Development

Phase 1: Application benchmarking and testing for existing state of art, benefits assessment

Data Layer – Streaming AMI, PMU, Distribution Models, OMS
Platform & Initial Distributed Comms
Simple Anomaly Detection – Learning Baseline Behavior

App 1: DR & DER Verification & Prediction	App 2: Distribution Incipient Failure	App 3: Topology & Parameter Estimation
PV Disaggregation Load response Dependency (FIDVR)	LTC failure analytics XFRMR Impedance detection	Load Identification Inverter Estimation Topology ID

Upper Supervisory Layer – OSISoft, OpenFMB integration

Integrated Multi Scale Data Analytics and Machine Learning

Approach 4 – Data and Industrial Support

Sentient ENERGY

PSL

PG&E

WATER | ENERGY | LIFE
CITY OF RIVERSIDE
PUBLIC UTILITIES

DUKE ENERGY

OSIsoft.

smarter grid solutions

OPAL-RT TECHNOLOGIES

sgip

Integrated Multi Scale Data Analytics and Machine Learning

Approach 5 – Metrics, testing and benefits

- Algorithms will be tested on reference (but real time) streaming data then evaluated against benefits framework for distributed multi-variate analysis
- Benefits framework will identify areas for development
  - Platform and distributed communications (latency, data quality)
  - Information prioritization (emergency vs normal ops)
  - New algorithm development (granularity of information, timeliness, ease of use)
  - Sensor fusion and flexibility of algorithms to data sources (can we use new data?)
- Metrics for success are tied directly to the use case and stakeholders and feed into phase 2

Integrated Multi Scale Data Analytics and Machine Learning

Key Milestones

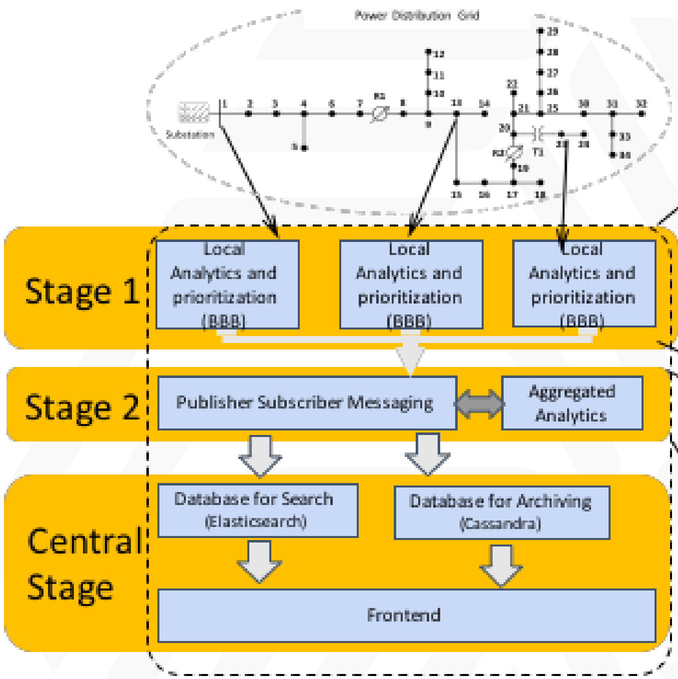
Milestone (FY16-FY18)	Status	Due Date
Task 1: White paper delivery and review	Draft Delivered & reviewed, publication in process	9/1/16 (complete)
Task 2: Workshop on white paper and use cases development, data gathering and use case specification complete	Workshop was delivered on Feb 9 216	2/1/17 (complete)
Task 3: Data collection, mapping of data to use case and platform access for team established	Data mapping presented at stakeholder review	2/1/17 (complete)
Task 4: Demonstration of selected use case with streaming data, with stakeholders, bench-top demonstration with real time streaming data validated	Use case selection in progress per 12/1/16 Benchtop data streaming platform demonstration in progress (uPMU and pqube data)	7/1/17
Task 5: Use cases developed within same framework, new algorithm development reviewed		9/30/17
Task 6: Framework proposed to integrate new data streams from sensors development tasks, benefits assessment		6/1/18

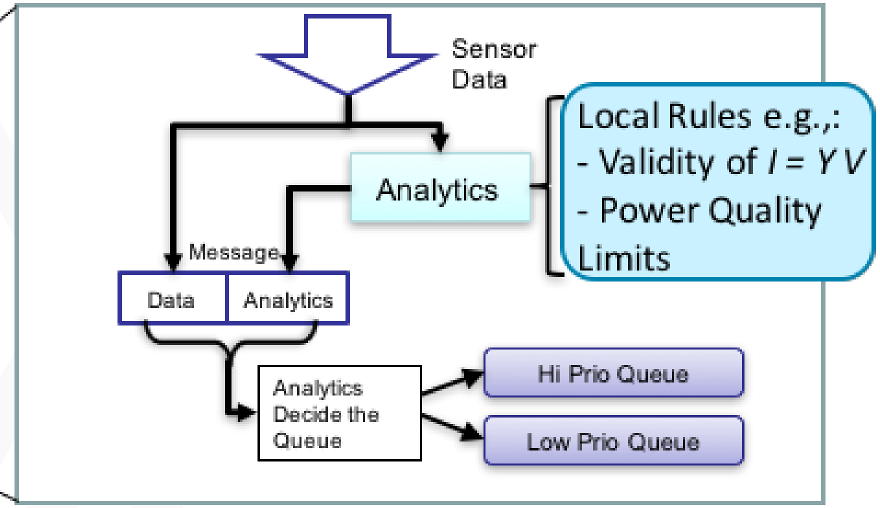
Integrated Multi Scale Data Analytics and Machine Learning Accomplishments to Date

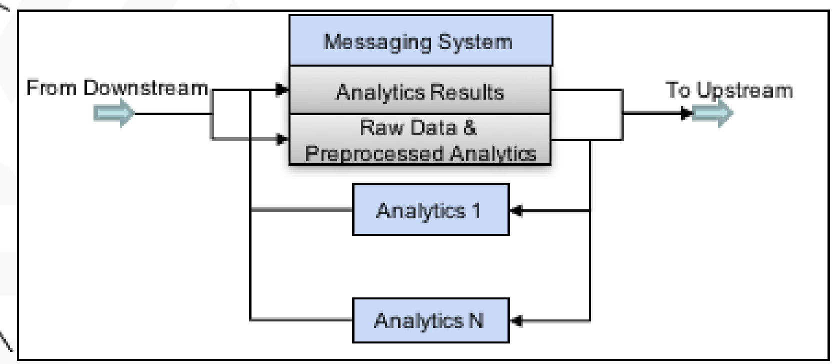
- Identified and reviewed with stakeholders, 3 high value use cases where new distributed ML techniques would have high impact on the building to grid interface
- Two white papers (in process of publishing)
- Structure for testing and benefits assessment of the existing state of the art is identified and initial application will be demonstrated in early July
- Coordinated with synergistic activities across programmatic boundaries

Integrated Multi Scale Data Analytics and Machine Learning

Accomplishments to Date: Platform Selection







- Archive the data & analytics results
- Frontend to the user

Platform Dev: Reinhard Gentz and Sean Peisert (LBNL)

Integrated Multi Scale Data Analytics and Machine Learning

Response to December 2016 Program Review

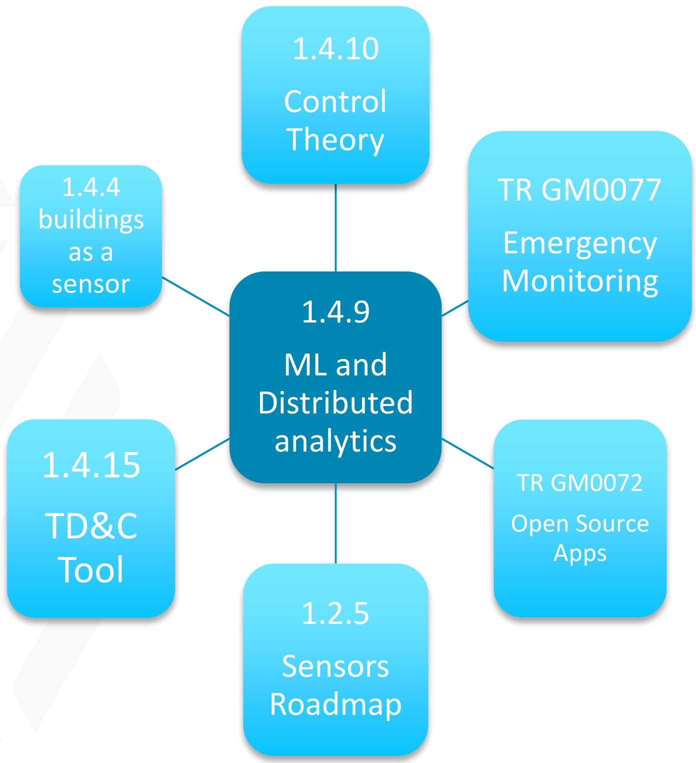
Recommendation	Response
Lab team should use the workshop planned in February to work with the stakeholder committee and identify the highest priority ML applications. Consult with the DOE program managers to select the best use cases moving forward	3 sets of use cases were presented at the stakeholder review meeting in February Questionnaire responses highlighted all 3 as being of importance, with incipient failure rating highest. These use cases were also highlighted as being of high importance to the S & M activities overall and have been integrated into the roadmapping work The team have presented this to the PM's and a strategy to review existing state of the art, and develop all applications concurrently within the multi-lab team.

Integrated Multi Scale Data Analytics and Machine Learning

Project Integration and Collaboration

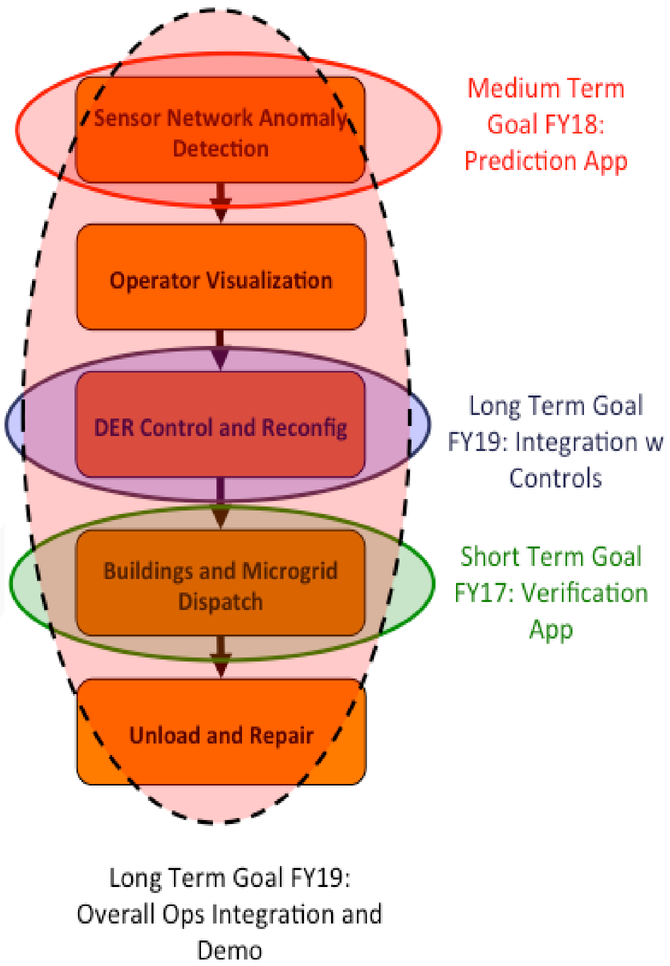
- GM0077: Anbient and Emergency Response – Scott Backhaus LANL
- GM0072: Load Model validation – Pavel Etingov LLNL
- 1.4.15: TDC test-bed development - Philip Top
- 1.4.10: Anomaly detection are precursors to control theory – Scott Backhaus
- 1.4.23: Threat Detection

- External project collaborations include ARPA-E uPMU for distribution, Sunshot ENERGISE and CEDS uPMU projects



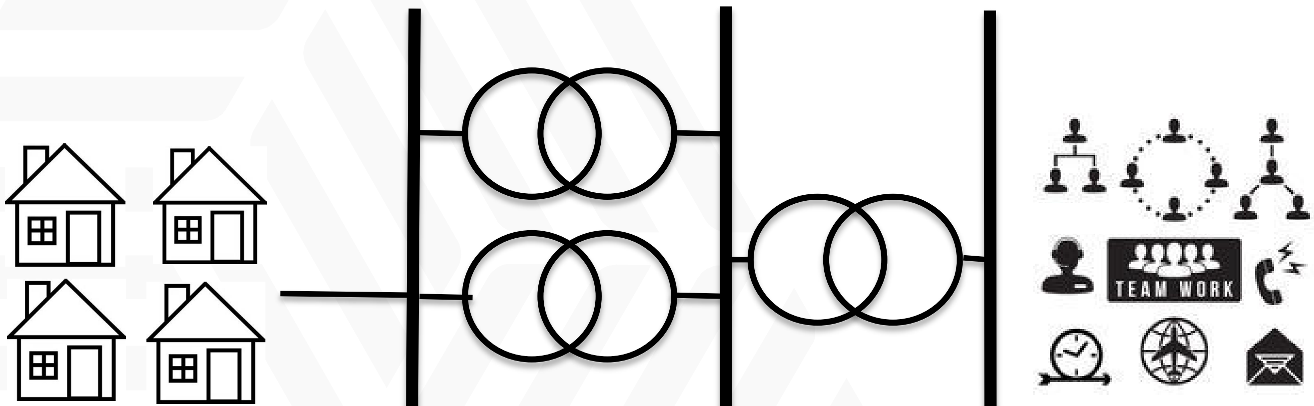
Integrated Multi Scale Data Analytics and Machine Learning
Next Steps and Future Plans

- **Primary goal of project is to develop architecture and analytics to transform data into actionable information – delivered to the right place, at the right time.**
- Next Steps: Complete first testing phase and report out benefits and requirements for development - July
- Outcome of phase 1 will include a map of activities required to meet final application development development
  - **Review demonstration with stakeholders at a workshop at LLNL**
  - **Conference papers for FY17**
- Phase 2 – new analytics techniques in each case study will be developed and implemented. Reference platform will be deployed at select locations and tested with enhanced features developed in phase 2
- Phase 3 – Selected analytics from the project will integrate with controls and upper layer hierarchy at BMS and DMS levels



Integrated Multi Scale Data Analytics and Machine Learning

Building to grid interface focus



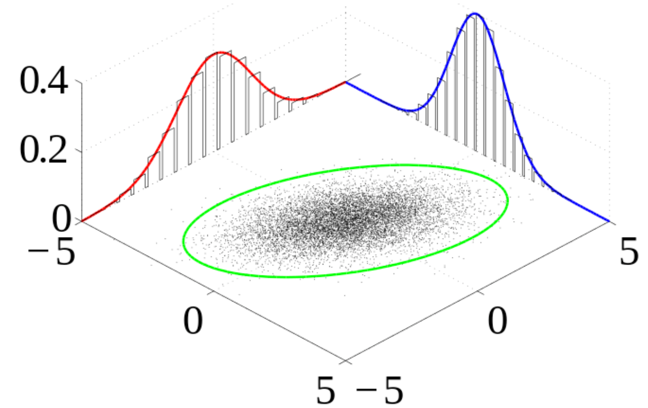
	LV – Customer Side	Distribution	Transmission	Ops and Planning
Timeframes	Hours, days, mins	us, s, mins	us, s, mins	Real time to 6 months+
Data	Green button, AMI, local, HVAC inverter	uPMU, Line sensor	SCADA, PMU	OMS, GIS, Models
Analytics Priorities	lighting, comfort, $$$	V, P, I management	Constraint management	CAIDI, SAIFI, MAIFI, safety
App 1 Impact	new markets, better economy	New services for granular automated management		
App 2 Impact	less outages, reduced cost of service	Preventative rather than reactive maintenance		
App 3 Impact	less outages, more efficient management	More accurate modeling, increased efficiency		

Operational intelligence

Predicting Ancillary Service Availability
Current Practice & Role of Machine Leaning

GRID MODERNIZATION INITIATIVE
U.S. Department of Energy

- Building Operator/Aggregator forecasts availability using deterministic techniques
  - Electric vehicles are scheduled and available capacity predicted and bid into markets
  - Solar PV production is forecasted
  - Load is forecasted as a function of temperature and time of day

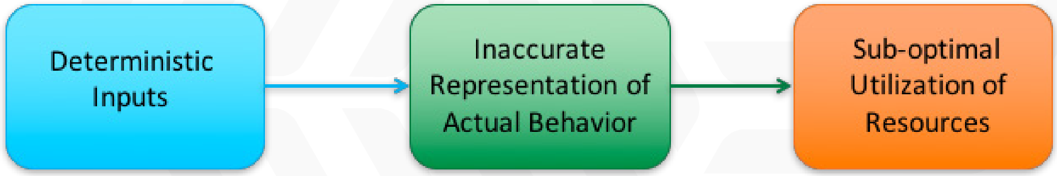


Machine learning can help automate and improve this process

Algorithms can understand complex intra-dependencies of processes, e.g., how does the scheduled electric vehicular fleet availability affect load

ML can better understand and account for stochastic behavior arising form occupant interaction and forecasts error and how these stochastic behaviors propagate through the system

- Following the formulation of forecasts the operator predicts the loads flexibility and its ability to provide ancillary services to the various markets and bids in this corresponding amount

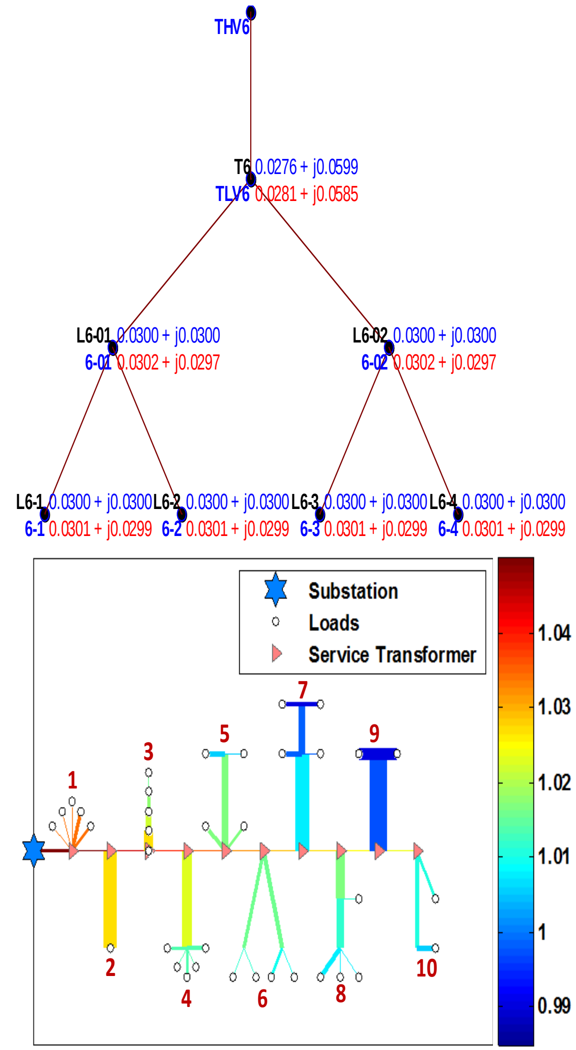


U.S. DEPARTMENT OF ENERGY

Grid Topology and Parameter Estimation New Approaches

GRID MODERNIZATION INITIATIVE
U.S. Department of Energy

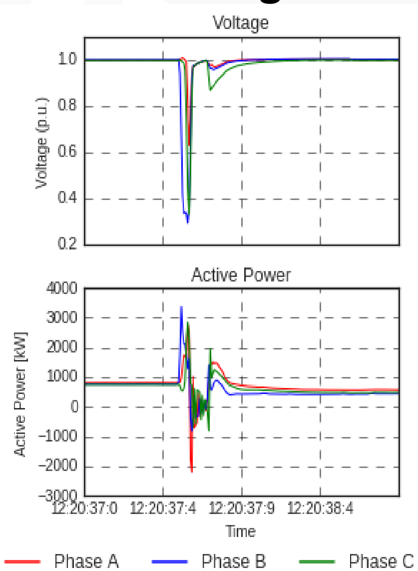
- New Machine Learning Approaches
  - Self-organizing maps for outlier and bad data detection
  - Random forest for topology identification
  - Robust regression for grid parameter estimation
- Sensing requirements
  - Historical power and voltage measurements from all buildings. Do not need high-resolution data. AMI data at 15-minute resolution, but for machine learning, several months of AMI data is required.
  - Meter accuracy is extremely important



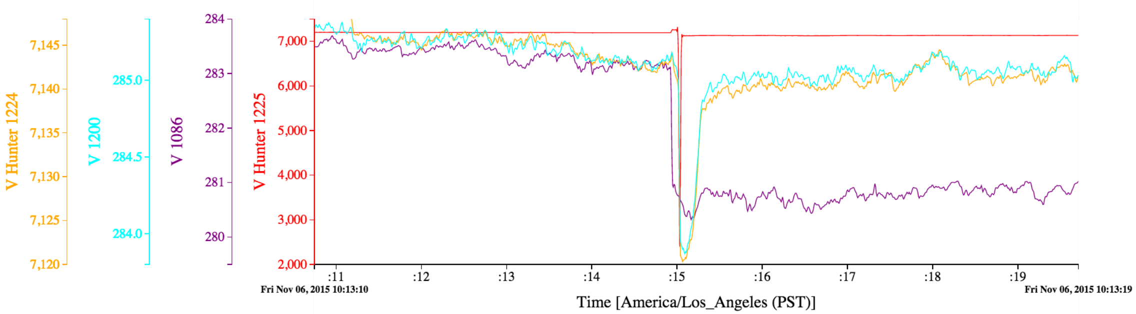
Case 2: Fault Analysis & Incipient Failure Problem and Background

- There are millions of distribution transformers and unmonitored equipment in the US
- Measurement of each individual device is economically and technically challenging
- Direct measurement approaches include Dissolved Gas Analysis, and
- Difficult to attribute anomalous behavior to a specific device or type of device

Bird Blowing a Fuse

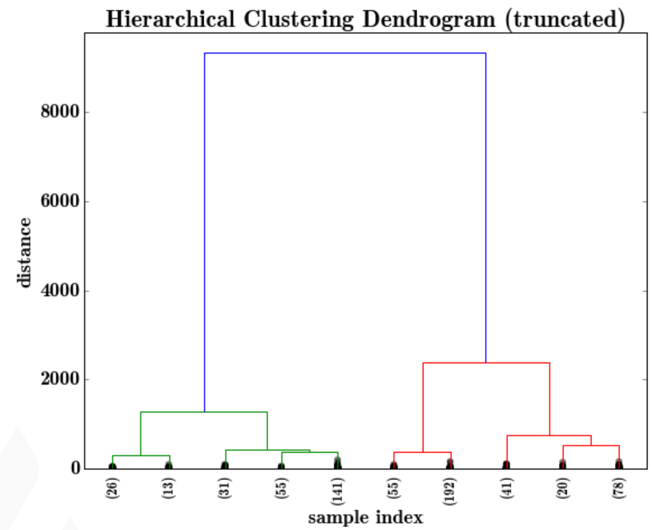


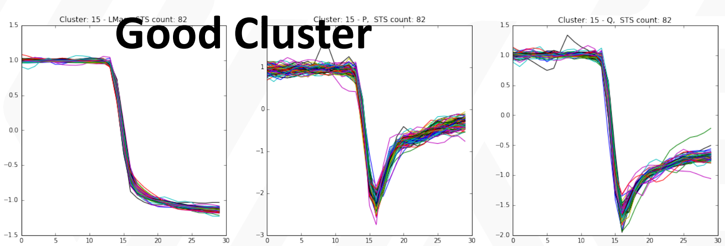
Tap Changer Oil Leak

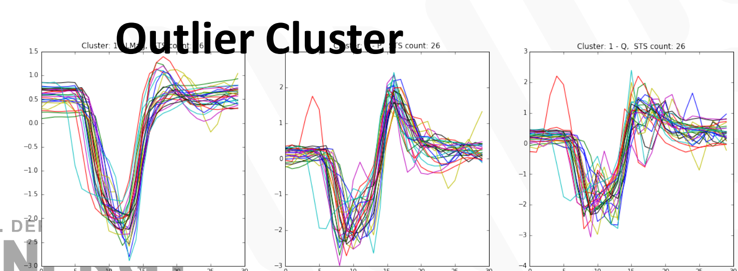


Equipment Incipient Failure New Approaches

- Hierarchical clustering allows for classification of power system phenomena via multi-dimensional clustering, both across phases and quantities
  - utilize derivative of phase angle as informative stream in clustering behavior
  - Can be applied to any time series behavior, utilize power flow modeling in relational analysis of potentially failing component







- Sensor needs
  - Phase angle and time series measurements
  - Relational and synchronized
  - Fused with power flow physics for locational analysis
  - Impact analysis and reconfiguration can utilize building data
  - Low accuracy as normal behavior is learned