IEEE PES Technical Webinar
Sponsored by IEEE PES Big Data subcommittee

Visualization and Analytics of Distribution Systems with Deep Penetration of Distributed Energy Resources (VADER)

Sila Kiliccote
Staff Scientist SLAC National Accelerator Laboratory
Managing Director, Precourt Institute for Energy @ Stanford

Outline of Presentation

- Problem statement
- Summary of the VADER Project
- Use Cases
- Data
- Architecture
- Data Analytics
- Impact

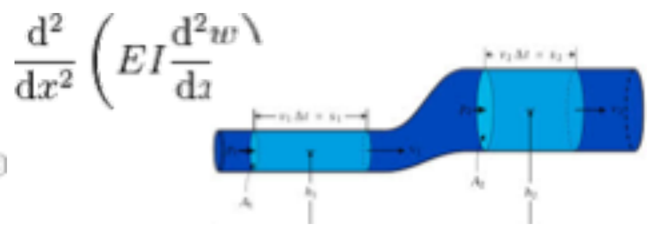
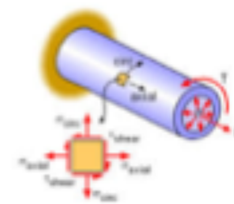
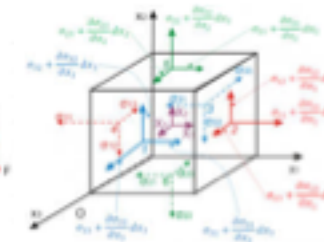
VADER is funded by the DOE SETO under SuNLaMP solicitation

Problem Statement

- More active devices that are not modeled or difficult to model.
- Utility unaware of small deployments and would like visibility behind the meter.
- Bi-directional power flow and over voltages.

Data-Driven Modeling for Power Systems

Physics-based models use basic equations of continuum mechanics, materials, heat transfer, power flow, …. that capture the phenomenon in a mathematical form



We don’t have ‘basic equations’ for social, medical, behavioral, economic and other complex phenomena.

Data-Driven Modeling has been extremely successful.

Take lots of data and fit the curve …

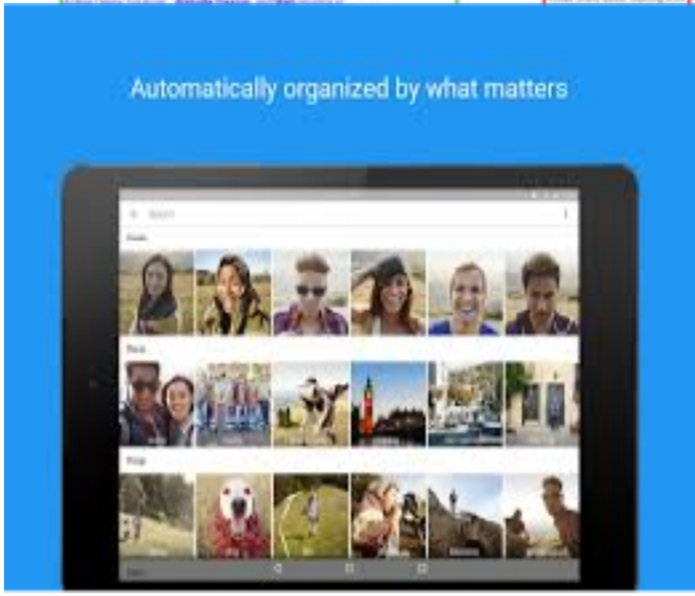
(No causal equations required)

Lots of data and compute power

Extremely successful in the last 10 years

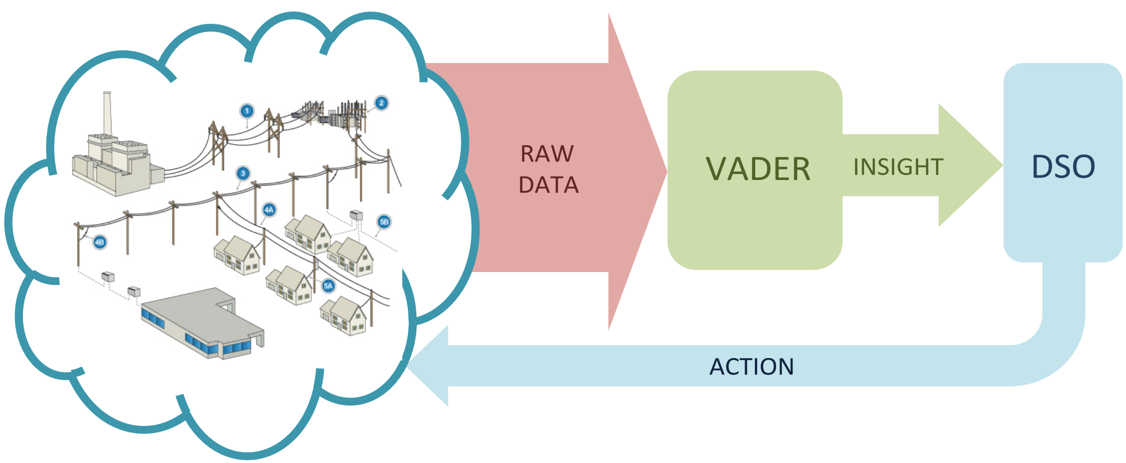
- Spell Correction
- Web search and advertising
- News feed
- Perception: Vision, speech

(Mostly web-ecosystem products)



Project Objectives

- Overall goal of the project: Understand the impact of technologies on the distribution system and how they can be used for planning and operations to increase PV penetration (reduce interconnection study costs and approval time) using a data-driven approach
  - build open-source tools to model and integrate a large number of data sources for distribution system planning and control,
  - verify capability of tools utilizing data from industry and utility partners,
  - validate the platform in a pilot testbed using HIL simulations and data from deployed hardware in the field



Key Metrics

- Scalable and dynamically adaptable platform to any PV penetration level
- Data accuracy, resolution, availability; system interoperability
- Open-source tools to model and integrate large number of data sources for distribution system control
- Reliable & real-time API access to PV, EV, AMI, SCADA, μ-PMU data
- Ease of data integration, processing and validation
- Robustness to missing data
- Integrate predictive analytics: state estimation, topology detection, scenario analysis; small predictive errors
- Demonstrate real-time visualization and monitoring
- Demonstrate the capability of integrating open source sensor placement algorithms

Approach to project

- Strategic Planning
- Data Collection and Integration
- Development of the Platform
- What now analytics
- What if analytics
- Network Analysis
  - Develop advanced topology identification capability
  - Develop sensor placement capability
  - Develop advanced state estimation capability

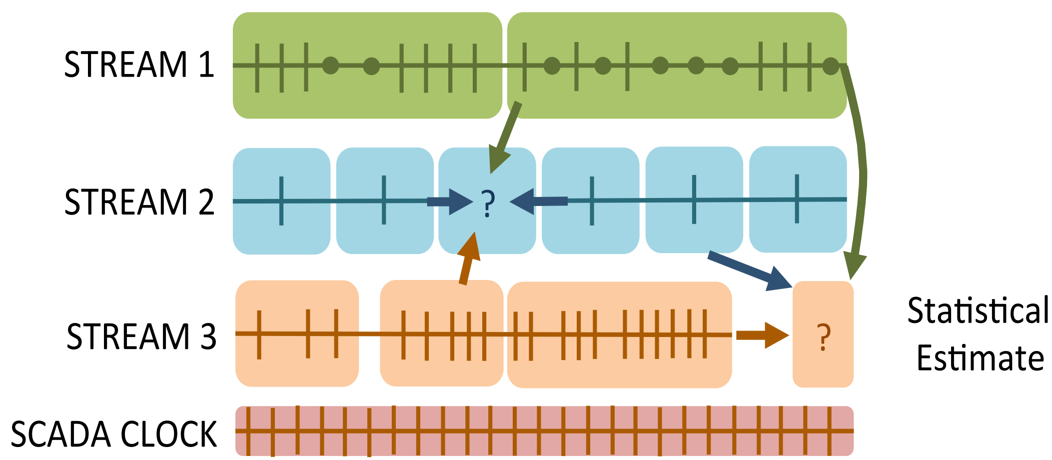
Project Innovations (1/3)d

- Basic Question: Planning and operations of a reliable, stable and efficient distribution system with high PV penetration (>100% of peak load) requires adequate monitoring and accurate prediction capability that allows scenario analysis and closed-loop control of the distribution system

- Vision of the Project: A unified data analytics platform that integrates *massive* and *heterogeneous* data streams for planning and granular real-time monitoring with analytics, visualization and control of distributed energy resources

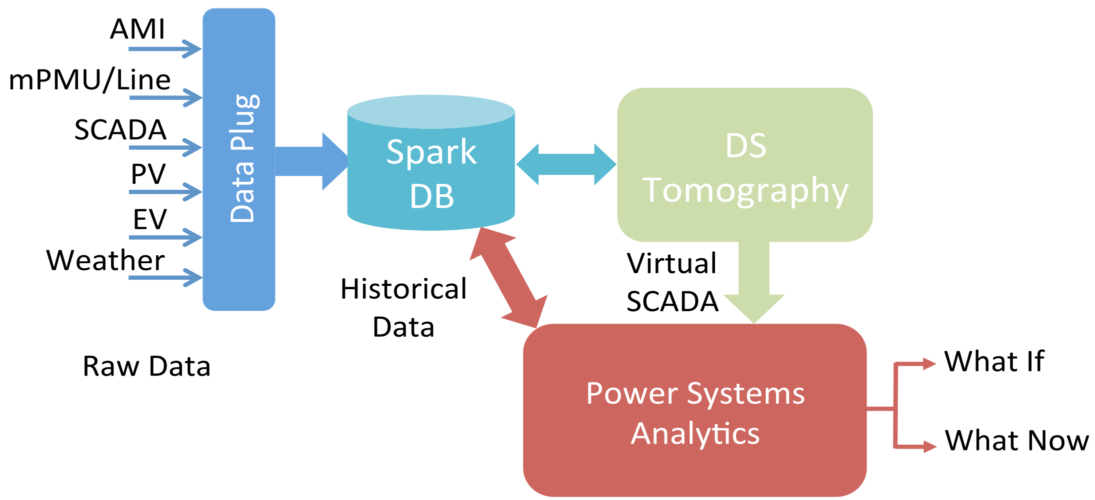
Project Innovations (2/3)

- Virtual SCADA system
  - Data Plug: interface with partner database APIs to stream data; advanced data management and validation tools
  - DS tomography module: integrates disparate, unreliable data into "virtual SCADA" data streams for power system analysis



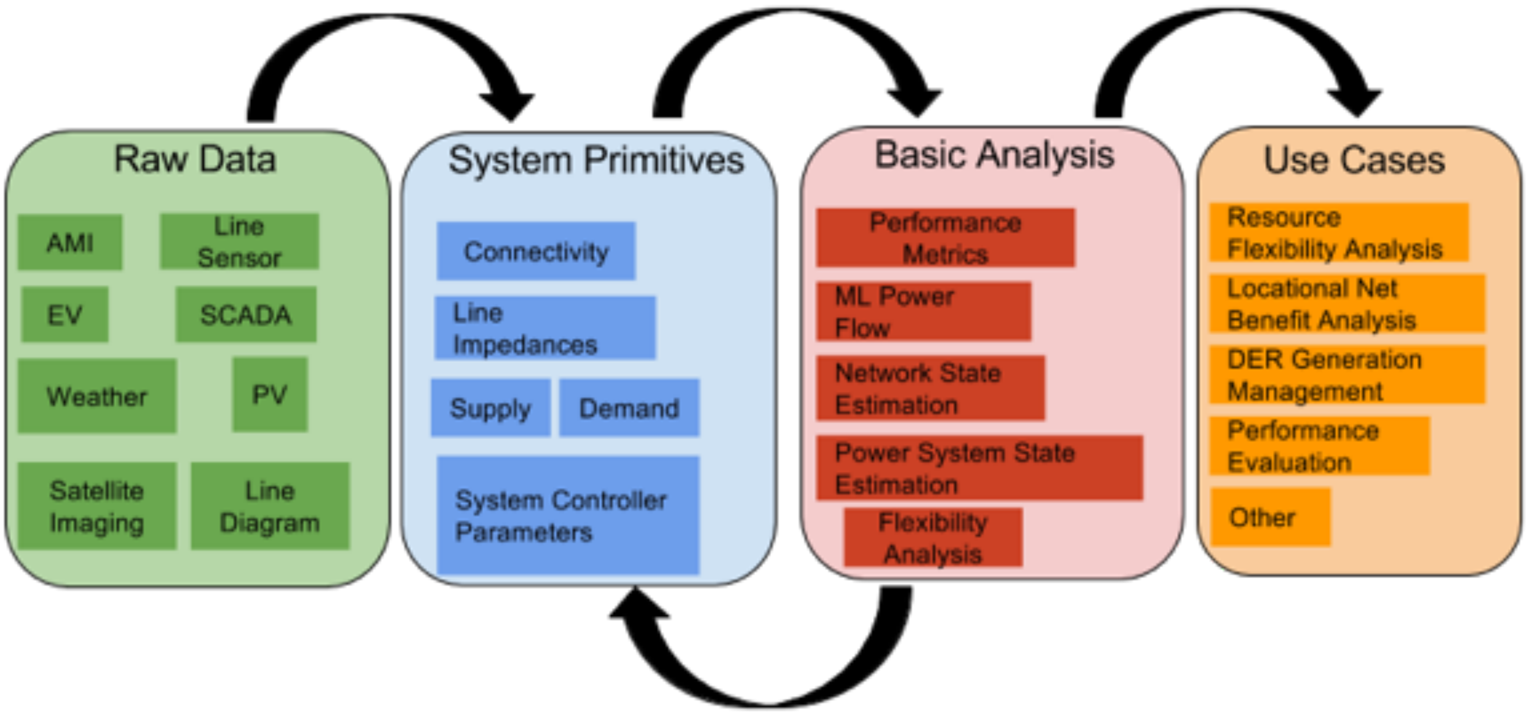
Project Innovations (3/3)

- VADER power system analytics tools
  - What-now analytics: advanced state estimation, situation awareness – outage detection, topology change
  - What-if analytics: analyze different scenarios of PV integration for planning; time-space analysis, location benefits

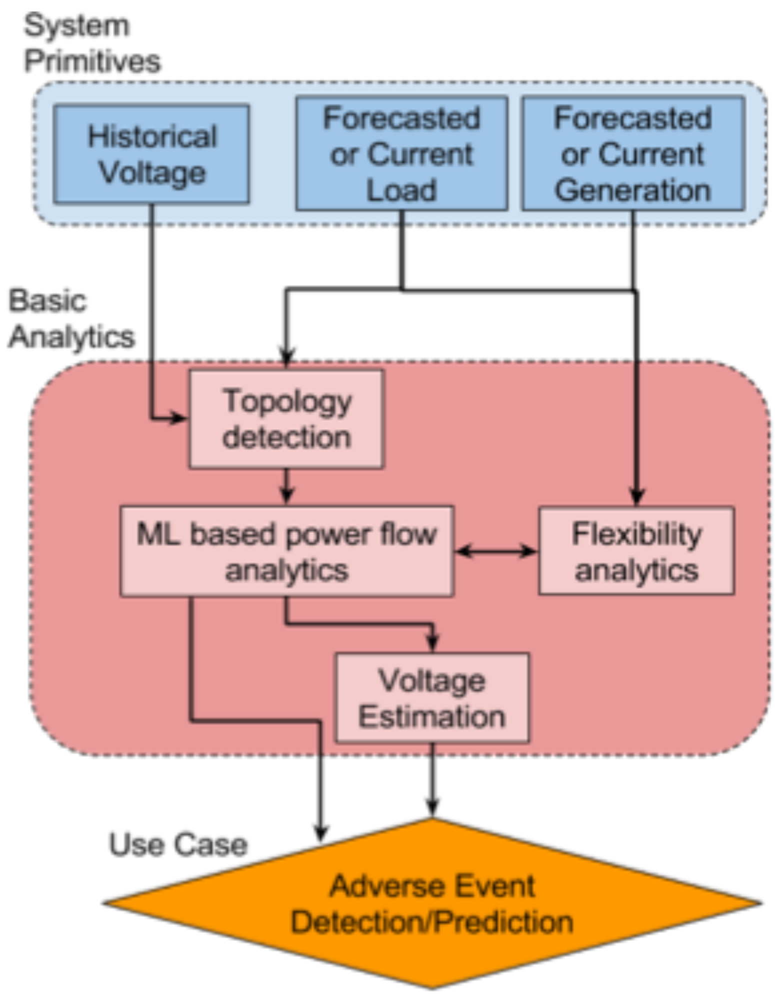


Use Cases

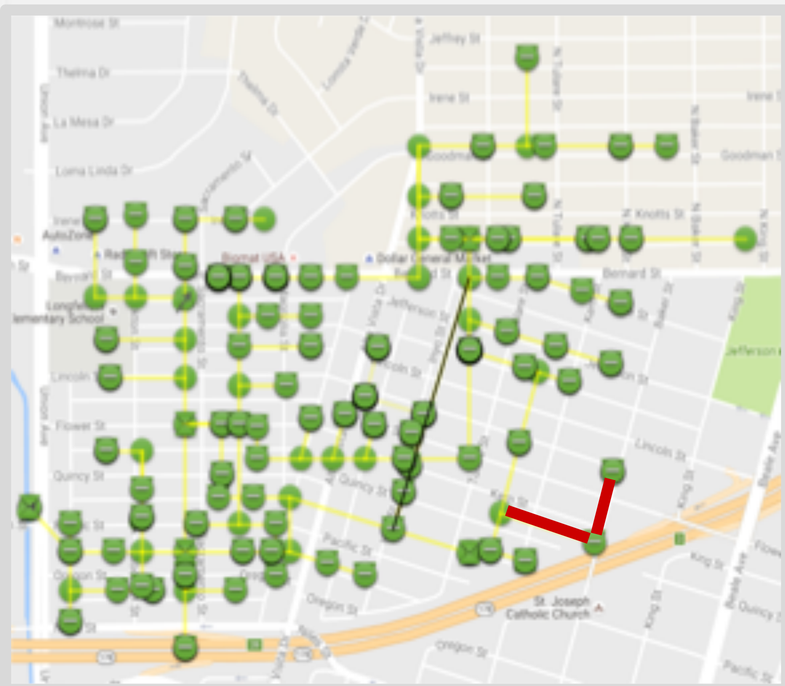
VADER Use Cases



Adverse event detection



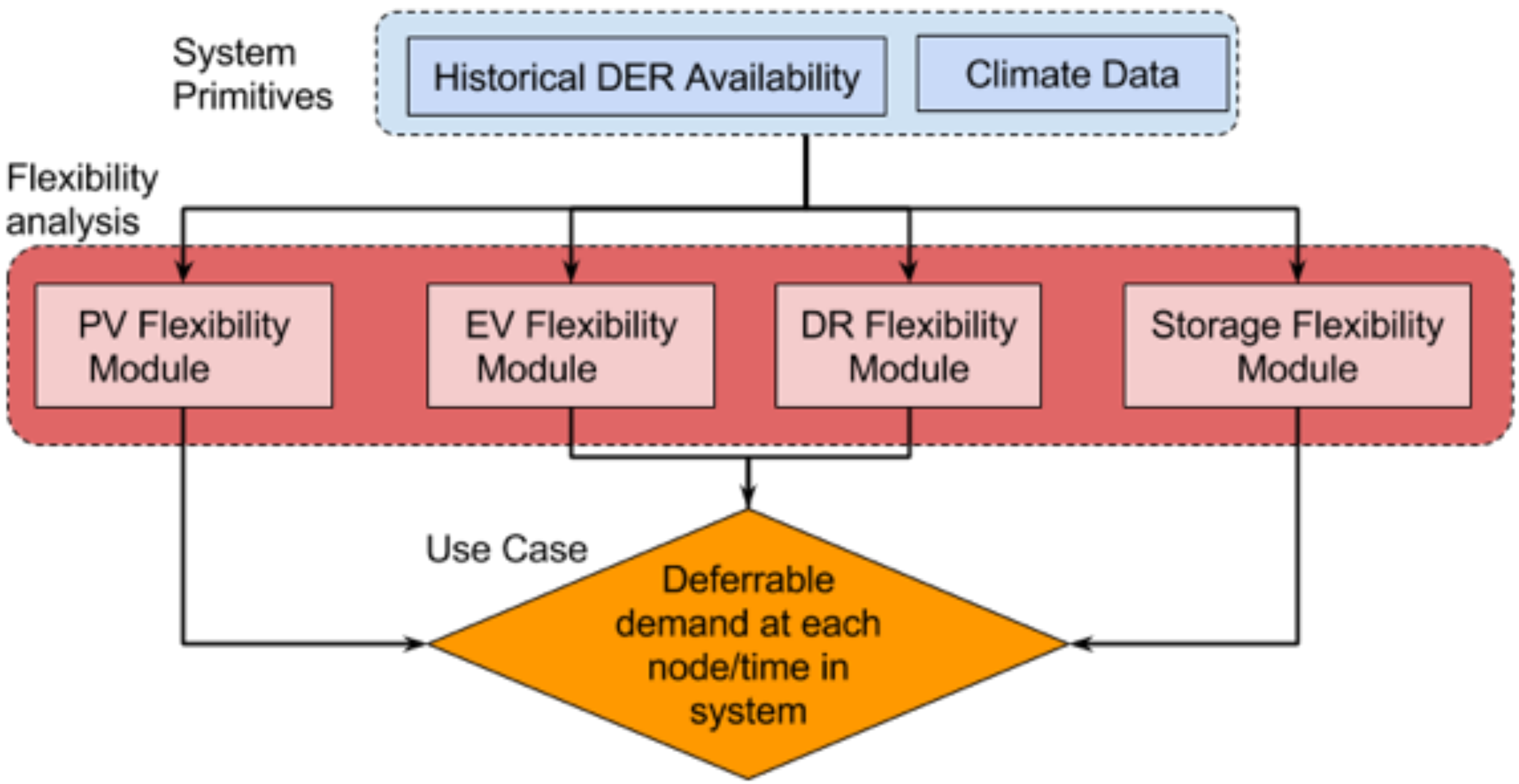
Adverse Event Detection Outputs



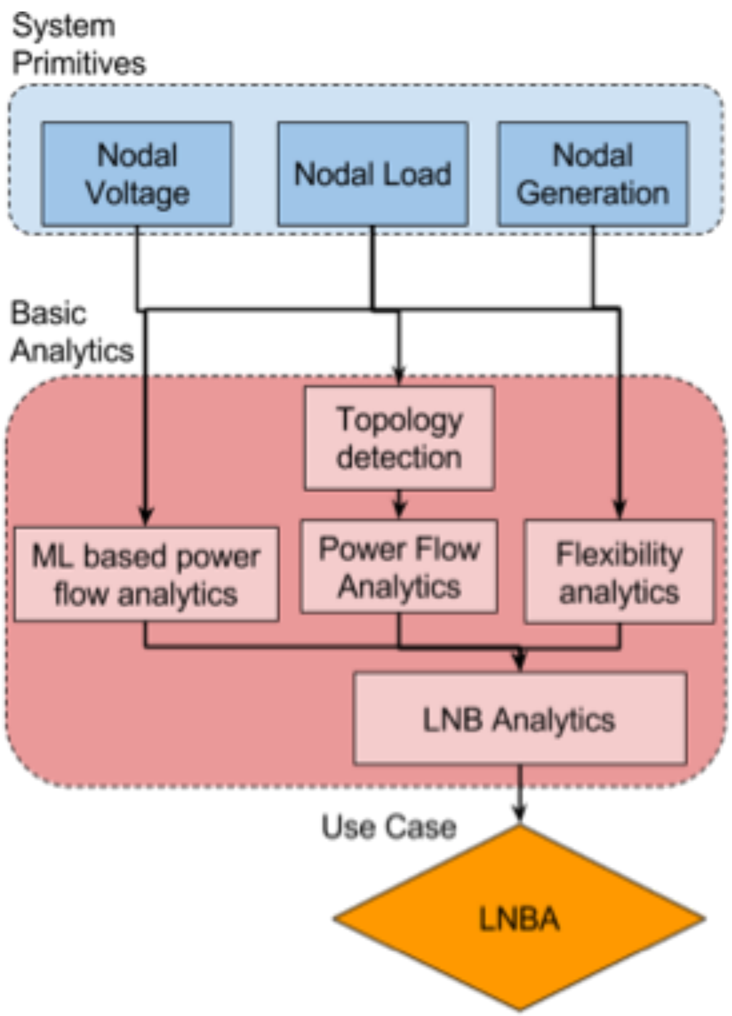
<Include>

- Interaction with the visualization
- Voltage and power flow statistics display
- Toggle before/after what-if scenario

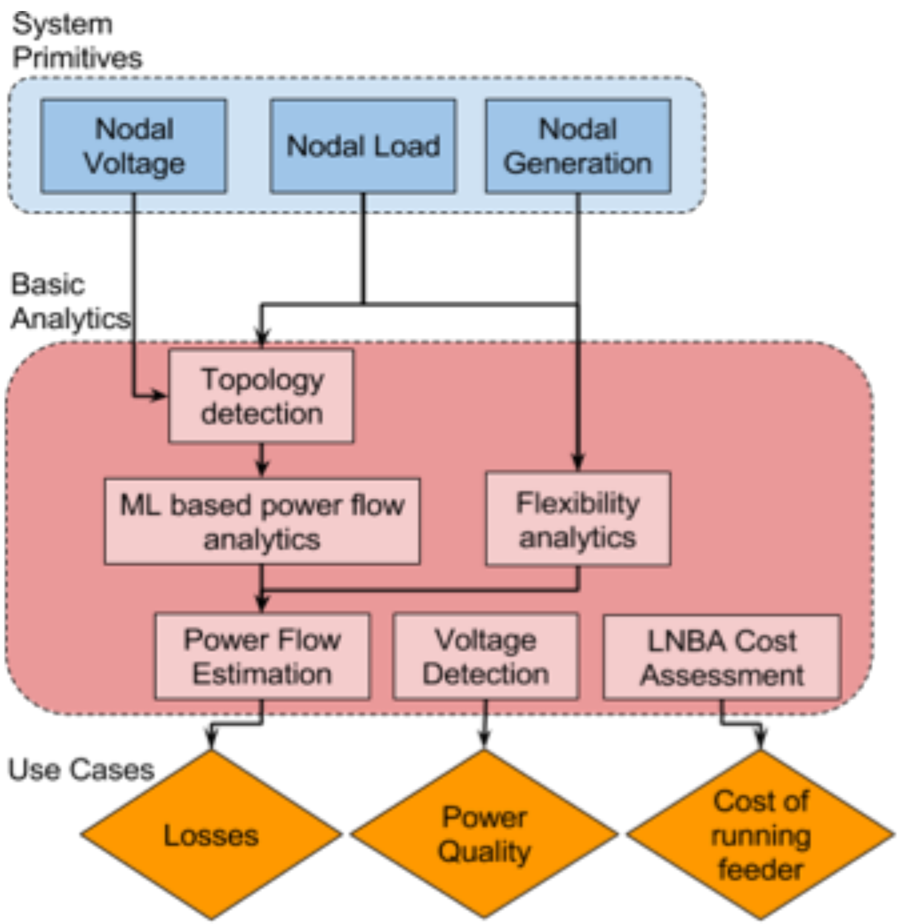
Resource Flexibility Analysis



Locational Net Benefits Assessment



Performance Evaluation of Distribution Systems



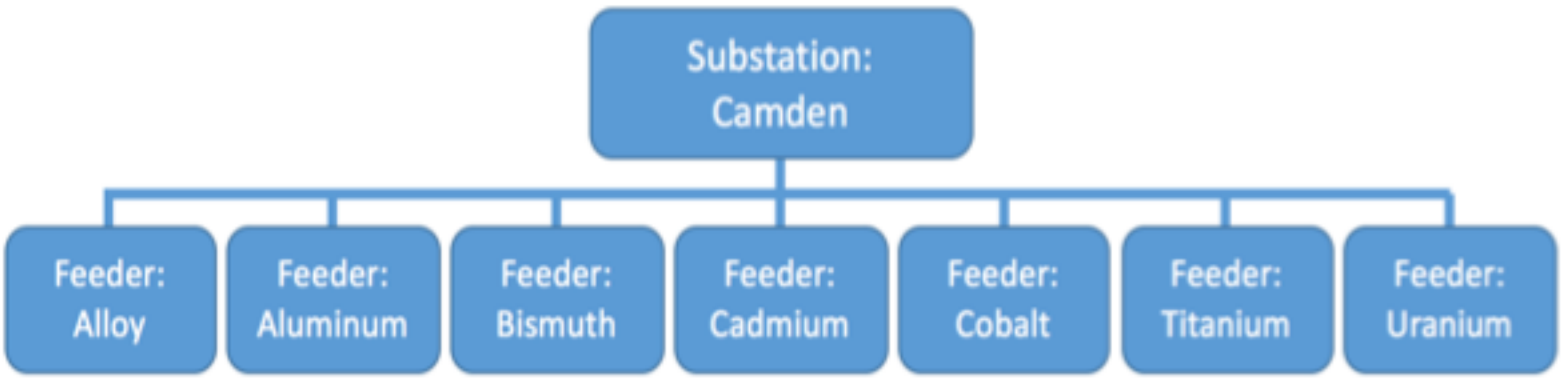
Lessons Learned

- Each utility has a different set of needs and priorities.
  - Some have meters on PVs
  - Some have sensors on all switches
  - All have common sets of needs…

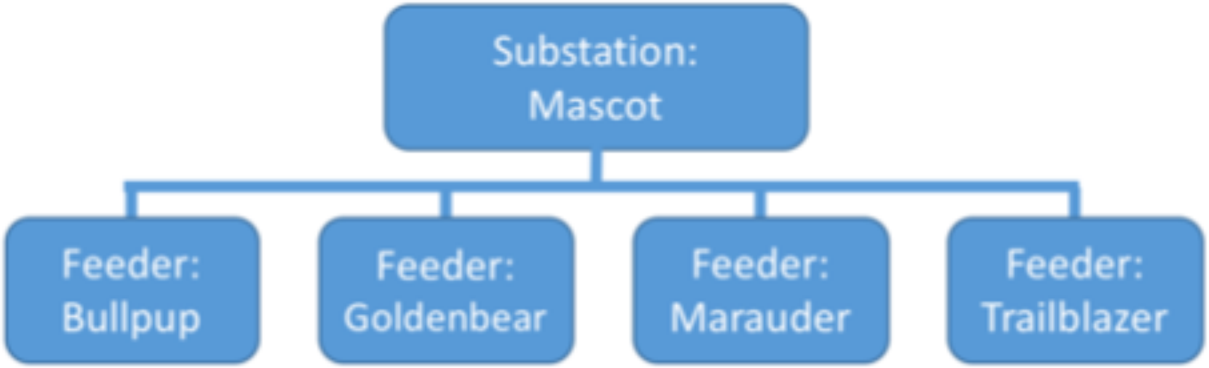
Data

Data

- Seven feeders of the Camden substation



- Four feeders of the Mascot substation



Data

- System data from the seven feeders of the Camden substation

Feeder Level	Alloy	Aluminum	Bismuth	Cadmium	Cobalt	Titanium	Uranium
Topology	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Google Earth Data	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cable Parameters	No	No	No	No	No	No	No
Root Node Real Power	No	Yes	No	No	No	Yes	No
Root Node Reactive Power	No	Yes	No	No	No	Yes	No
Root Node Voltage Magnitude	No	Yes	No	No	No	Yes	No
Root Node Voltage Phase Angle	No	Yes	No	No	No	Yes	No
All buses loads	Yes	Yes	Yes	No	No	No	No
All buses Reactive Power	No	No	No	No	No	No	No
All buses Voltage Magnitude	Yes	Yes	Yes	No	No	No	No
All buses Voltage Phase Angle	No	No	No	No	No	No	No

Data

- System data from the four feeders of the Mascot substation

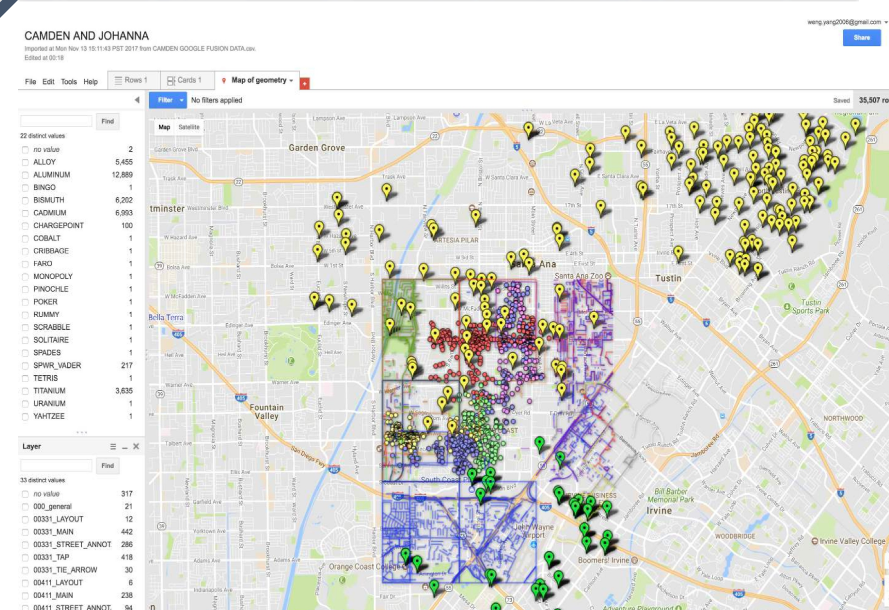
Feeder Level	Bullpup	Goldenbear	Marauder	Trailblazer
Topology	Yes	Yes	Yes	Yes
Google Earth Data	No	No	No	No
Cable Parameters	No	No	No	No
Root Node Real Power	No	No	No	No
Root Node Reactive Power	No	No	No	No
Root Node Voltage Magnitude	No	No	No	No
Root Node Voltage Phase Angle	No	No	No	No
All buses loads	Yes	Yes	Yes	Yes
All buses Reactive Power	No	No	No	No
All buses Voltage Magnitude	Yes	Yes	Yes	No
All buses Voltage Phase Angle	No	No	No	No

Lessons Learned

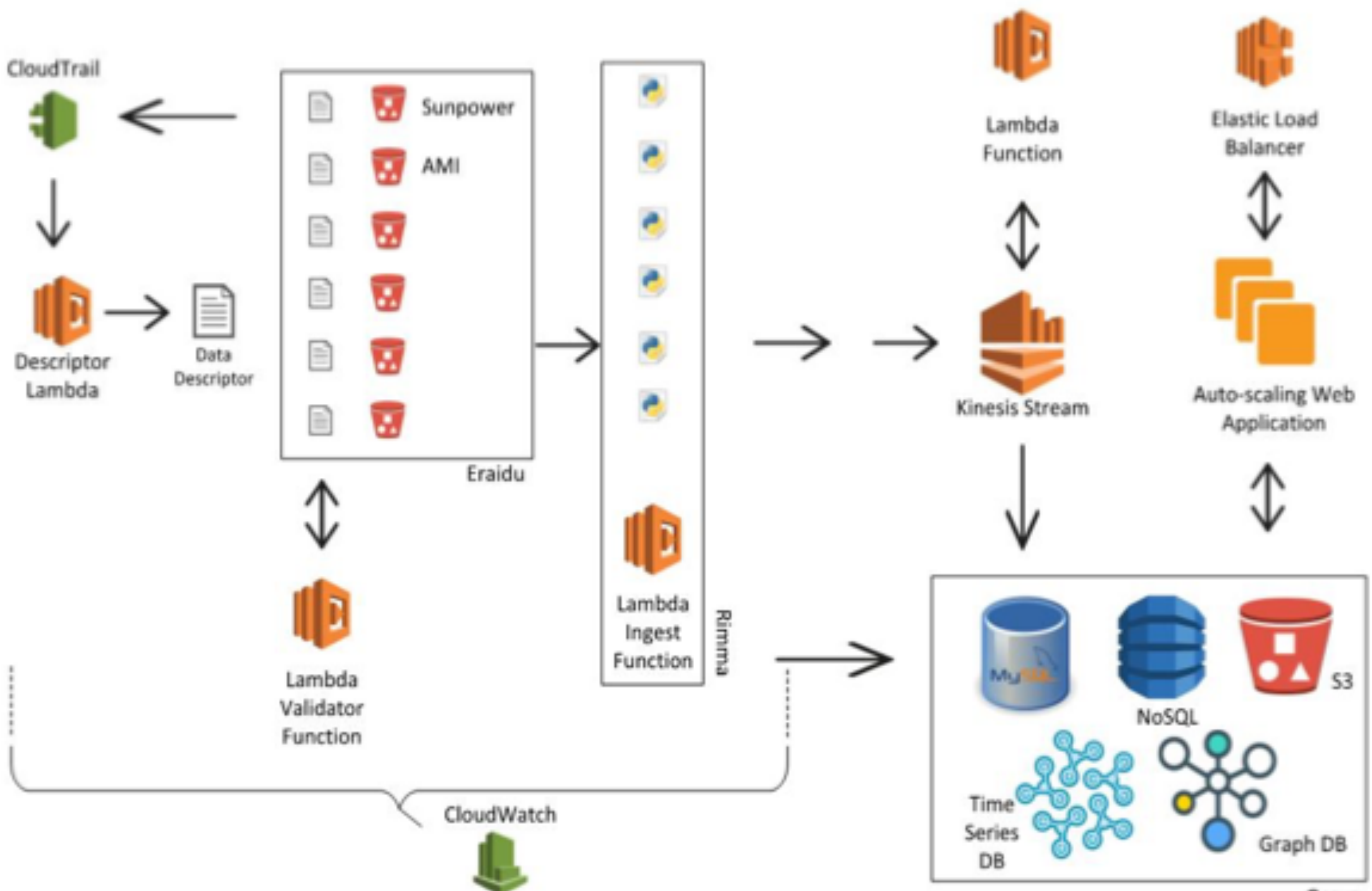
- Data comes in different shapes and sizes
- Developing schemas is time consuming major effort
- Open schemas to support utility data integration is needed

Architecture

Platform evolution



VADER
System Architecture



Lessons learned

- Single instance vs. multi tenant
- Synchronization (HE, HB, daylight savings)
- Raw data vs. processed data
- Encryption
- User and developer interface

Data Analytics

Utilization of Data for Power System Analytics Tools (1/2)

Power System Analytics	Types of Data
(1) Topology reconstruction	(1.1) Hourly bus voltage magnitudes * Two weeks of data for training * Extensive testing & validation using one year of data
(2) Switch configuration detection	(2.1) 1-minute resolution of voltage data * Data for several hours
(3) Line parameter estimation	(3.1) Phase angles from μ-PMU data in addition to data type (1.1) (3.2) Active (*P*) and reactive power (*Q*), if available
(4) Outage detection	(4.1) Bus voltage magnitude and phase angle from μ-PMU
(5) Machine learning-based power flow	(5.1) Data synchronization between utility and third parties (e.g. PV data from SunPower), data plug module (5.2) *P* & *Q* at each bus (5.3) SunPower voltage magnitude and its *P* at solar locations will improve estimates

Utilization of Data for Power System Analytics Tools (2/2)

Power System Analytics	Types of Data
(6) Solar disaggregation	(6.1) Net load measurements at the point of disaggregation (three scenarios): At substation (SCADA ~4 sec sampling rate) / transformer (aggregated from AMI downstream) / AMI meters (15-minute or faster) (6.2) Outside temperature from the region of interest (6.3) solar proxy; data from irradiance sensors and/or active power measurements, typically 1-2 minute sampling rate (6.4) Reactive power, if available, at substation / transformer / AMI at same sampling rate as load
(7) Customer load forecasting	(7.1) Hourly smart meter active power * Two weeks of data for training
(8) Clear sky solar prediction	(8.1) measured output power of PV system

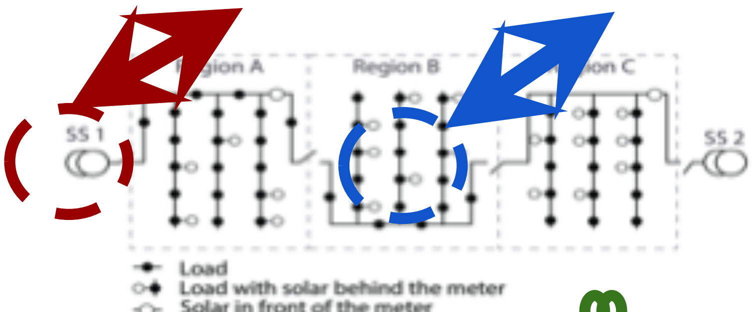
Solar Disaggregation: gain visibility into behind-the-meter solar (Emre Kara, Michaelangelo Tabone)

Disaggregate solar generation from meter readings of net load

<u>2 measurements of net load in distribution systems</u>

Real-time SCADA measurements
Typical 4 seconds sampling

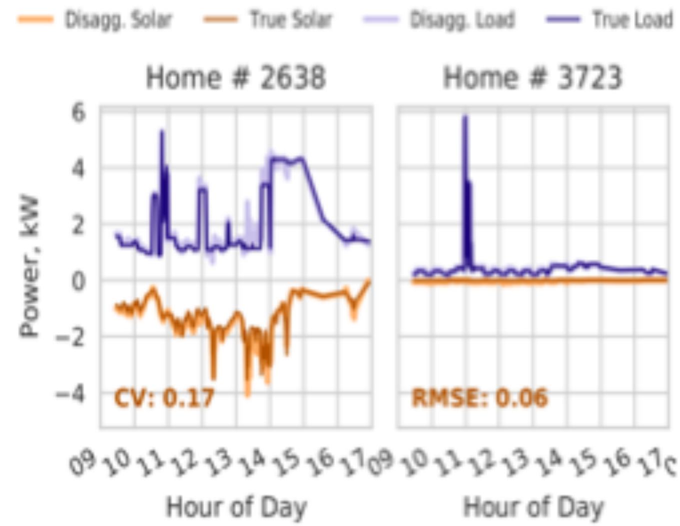
AMI: overnight updates
1-min to hourly sampling rate



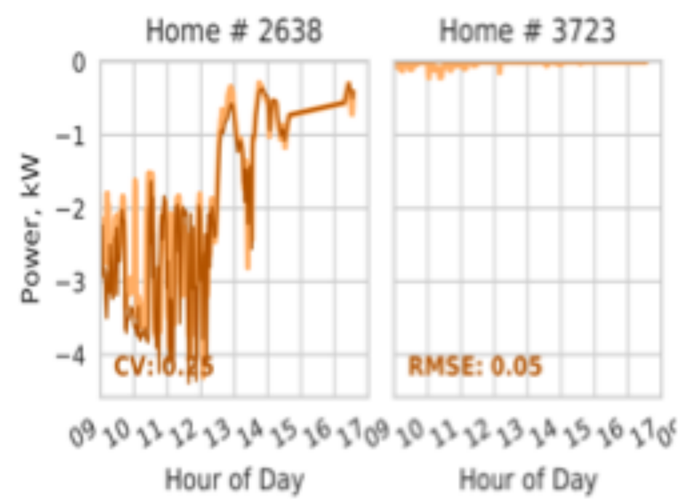
Initial work at LBNL doing feeder level solar-disaggregation has received an R&D 100 award.

Validation set: submetered solar and load from Pecan Street

Day Ahead Model Training
AMI + SCADA

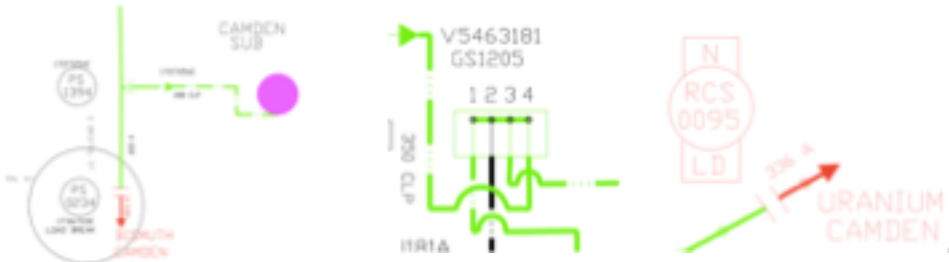


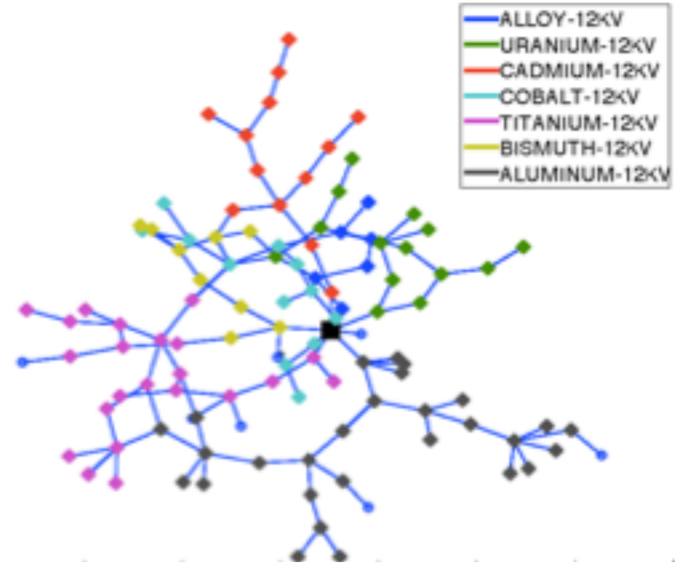
Streaming Analytics
SCADA



SCE Radial Switch Configuration Detection

Camden Substation





Inter/Intra Feeder:

Underground, Pole Top, Remote Controlled

Network Summary

112 Aggregated Loads with 1, 2, 3 phase loads.
123 Switches to Monitor

5.76057e+09 Possible Radial Configurations

Theory Predicts:

AMI + 12 Line Measurements vs. 123 SCADA Sensors

Current Work:

Extending Algorithms for lossy/3-phase networks.

Machine Learning-based Power Flow (Ram Rajagopal)

Availability of topology line parameters

- Traditional state estimation method: require line connectivity and parameters information
- ML method: *no need for line Information*

Ability to handle missing measurements

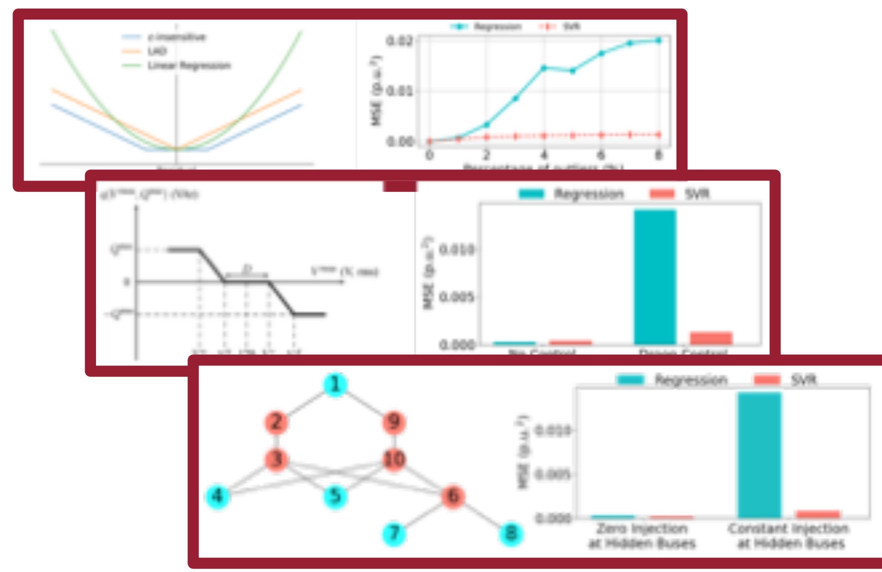
- Traditional Method: No. It needs the whole system to be observable.
- ML Method: Yes. It only *builds correlation between available data at available time slots.*

Ability to conduct voltage forecasting / power flow

- Traditional Method: No. It is static state estimation.
- ML Method: Yes. It only *builds correlation between voltages and power*, forecast power, and recover voltage based on the relationship.

Machine Learning Based Power Flow -
How does it work and how does it compare

- Practical Advantages of [illegible] Flow
  - Equivalence to physical model
  - Robustness against outliers
  - Capability of modeling 3rd party controllers
  - Flexibility for partially observed systems model construction
  - Capability of inverse mapping: P, Q to voltage mapping



Yu, Jiafan, Yang Weng, and Ram Rajagopal. "Mapping Rule Estimation for Power Flow Analysis in Distribution Grids." *arXiv preprint arXiv:1702.07948*(2017).

Lessons learned

- Working with utility data is not a streamlined process
- There are no lack of ideas in data analytics – if you build it they come.
- No rigorous way to test analytics performance
  - data is often not available to be shared
  - No common data sets for folks to compare performance
  - No platforms that allow analytics comparison

Impact

Industry Engagement - Workshops and Learning Lab

Two workshops hosted at SLAC

Goal: to receive critical review

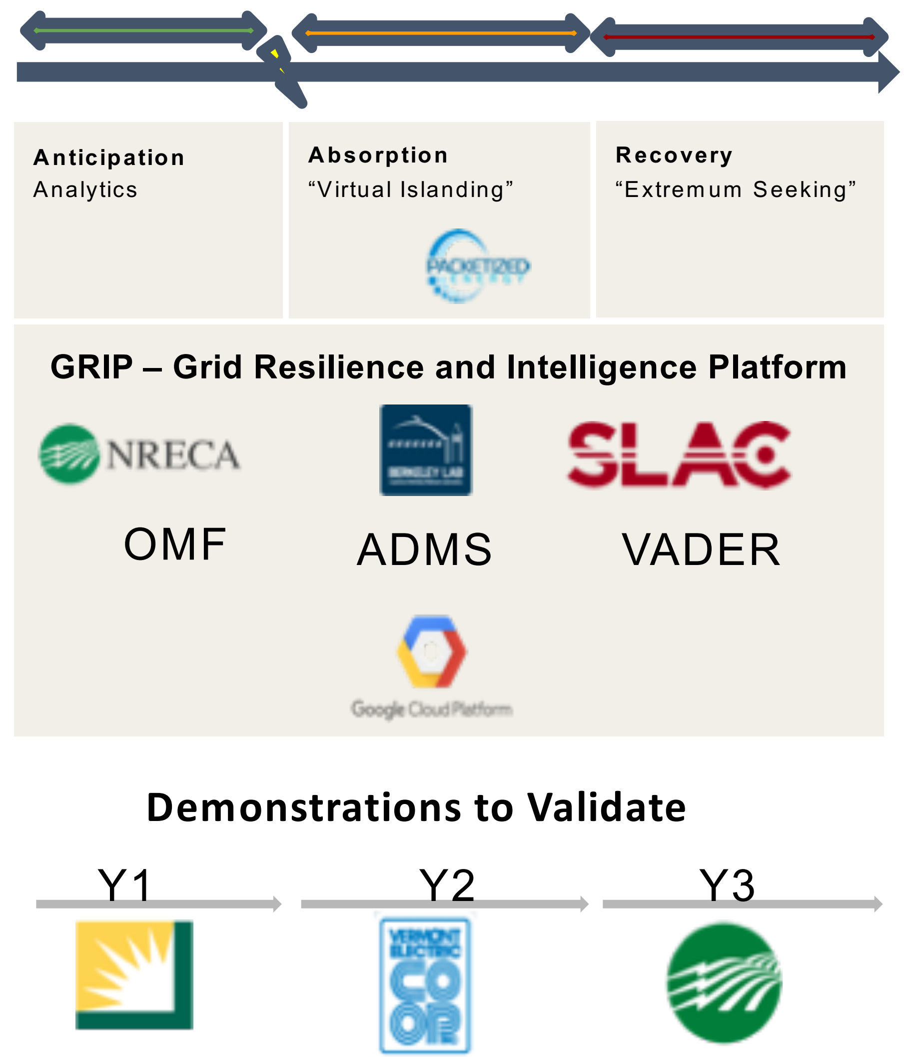
Two VADER Learning Labs hosted:

- End of March 2017 @ SLAC: industry participation
- End of May 2017 @ California Energy Commission: CEC staff participation

Goal: Increase awareness to drive adoption

GRIP - Grid Resilience and Intelligence Platform

- **Anticipation** building on current analytics capabilities have been discussed with NRECA and SCE. A prioritized list will be developed with broader stakeholder input.
- **Absorption** will focus on demonstrating virtual islanding.
- **Recovery** validations will focus on DER control without communications.



Publications

[1] Yizheng Liao, Yang Weng, and Ram Rajagopal, "Urban Distribution Grid Topology Reconstruction via Lasso", *Proceedings of IEEE Power and Energy Society General Meeting*, 17-21 July, 2016.

[2] Yizheng Liao, Yang Weng, Chin-Woo Tan, and Ram Rajagopal, "Urban Distribution Grid Line Outage Detection",*Proceedings of IEEE Conference on Probabilistic Methods Applied to Power Systems*, Beijing, China, 16-20 October, 2016. (Best Conference Paper Award)

[3] Emre C. Kara, et al., "Estimating Behind-the-meter Solar Generation with Existing Measurement Infrastructure",*Buildsys'16 ACM International Conference on Systems for Energy-Efficient Built Environments*, November 2016.

[4] Emre C Kara, et al., "Towards real-time estimation of solar generation from micro-synchrophasor measurements",*arXiv preprint arXiv:1607.02919* (2016).

[5] Junjie Qin, Insoon Young, and Ram Rajagopal, "Submodularity of Energy Storage Placement in Power Network", *Proceedings of IEEE Conference on Control and Decision*, 12-14 December, 2016.

[6] Souhaib Ben Taieb, Jiafan Yu, Mateus Neves Barreto, and Ram Rajagopal, "Regularization in Hierarchical Time Series Forecasting with Application to Electricity Smart Meter Data", *Proceedings of AAAI conference on Artificial Intelligence*, 4-9 February, 2017.

[7] Jiafan Yu, Junjie Qin, and Ram Rajagopal, “On Certainty Equivalence of Demand Charge Reduction Using Storage”, *Proceedings of American Control Conference*, Seattle, WA, 24-26 May, 2017.

[8] Bennet Meyers and Mark Mikofski, “Accurate Modeling of Partially Shaded PV Arrays”*, Proceedings of Photovoltaic Specialists Conference* (PVSC-44), Washington, DC, 25-30 June, 2017.

Publications cont.

[9] Jiafan Yu, Yang Weng, and Ram Rajagopal, “Data-Driven Joint Topology and Line Parameter Estimation for Renewable Integration”, *Proceedings of IEEE Power and Energy Society General Meeting,* Chicago, IL, 16-20 July, 2017.

[10] Jiafan Yu, Yang Weng, and Ram Rajagopal, “Robust Mapping Rule Estimation for Power Flow Analysis in Distribution Grids”, North American Power Symposium, Morgantown, WV, 17-19 September, 2017.

[11] M. Malik et al. “A Common Data Architecture for Energy Data Analytics”, IEEE SmartGridComm

[12] Nikolay Laptev, Jiafan Yu, and Ram Rajagopal, “Deepcast: Universal Time Series Forecaster”, *International Conference on Learning Representations*, 2017.

[13] Raffi Sevlian and Ram Rajagopal, "Distribution System Topology Detection Using Consumer Load and Line Flow Measurements." *arXiv preprint arXiv:1503.07224* (2017).

[14] Yizheng Liao, Yang Weng, and Ram Rajagopal, “Distributed Energy Resources Topology Identification via Graphical modeling”, *IEEE Transactions on Power Systems,* 2017 (accepted for publication).

[15] Yizheng Liao, Yang Weng, Guangyi Liu, and Ram Rajagopal, “Urban MV and LV Distribution Grid Topology via Group Lasso”, *IEEE Transactions on Power Systems,* 2017 (under review). http://web.stanford.edu/~yzliao/pub/TPS_info.pdf.

[16] Jiafan Yu, Yang Weng, and Ram Rajagopal, “PaToPa: A Data-Driven Parameter and Topology Joint Estimation Framework in Distribution Grids”, *IEEE Transactions on Power Systems* (under review).

VADER Accomplishments

- Initial set of analytics developed and tested with IEEE-123 Bus Model (GridLab-D integration) and some validated with actual data
  - Machine Learning-based Power Flow
  - Switch Detection
  - Solar Disaggregation
  - Forecasting
  - Topology detection
  - Statistical Clear Sky
  - PV Power Intraday Forecasting
- Platform demonstration with historical data
- Platform transition to more scalable implementation with real-time data
- Held VADER Workshops and Labs
- Started applying Southern California Edison's data
  - Solar Disaggregation
  - Switch Detection
- Expanded machine learning-based Power Flow to three-phase systems.
- Continue to improve analytics

Next steps

- Continue to validate existing analytics
- Implement LNBA as a use case
- Post open source code and documentation on Github (and make it easy to find)
- Continue to use the platform for other research projects
  - Baseline study
  - Smart Charging Infrastructure Planning Tool

VADER Team

SLAC
Emre Kara, David Chassin, Mayank Malik, Raffi Sevlian, Supriya Premkumar, Alyona Ivanova, Bennet Meyers, Berk Serbetcioglu
Stanford University

Ram Rajagopal, Chin-Woo Tan, Michaelangelo Tabone, Mark Chen, Yizheng Liao, Jiafan Yu, Yang Weng, Siobhan Powell

+ 15 Carnegie Mellon University INI Practicum Students

Thank you.

Sila Kiliccote
silak@stanford.edu

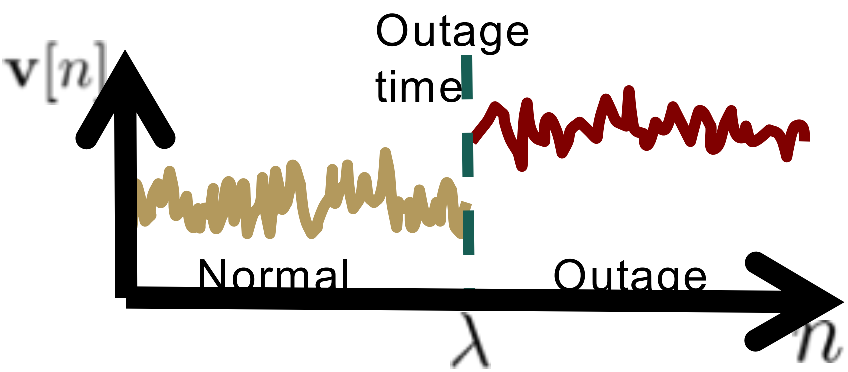
Extra Sildes

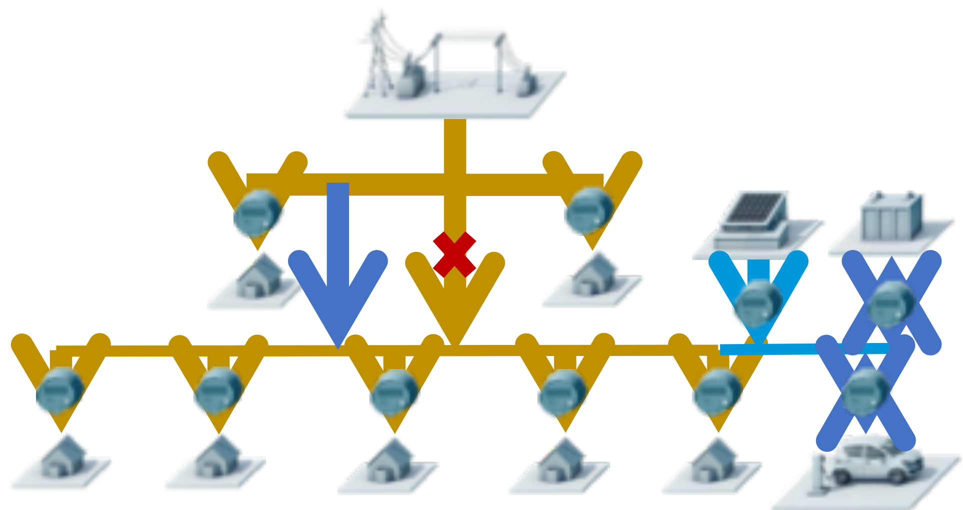
CIS/OMS Sensor Data Plug Validation (cont'd)

- Problem formulation

Time-series voltage magnitude

$$\mathbf{v}^{1:N} = \{\mathbf{v}[1], \mathbf{v}[2], \cdots, \mathbf{v}[N]\}$$



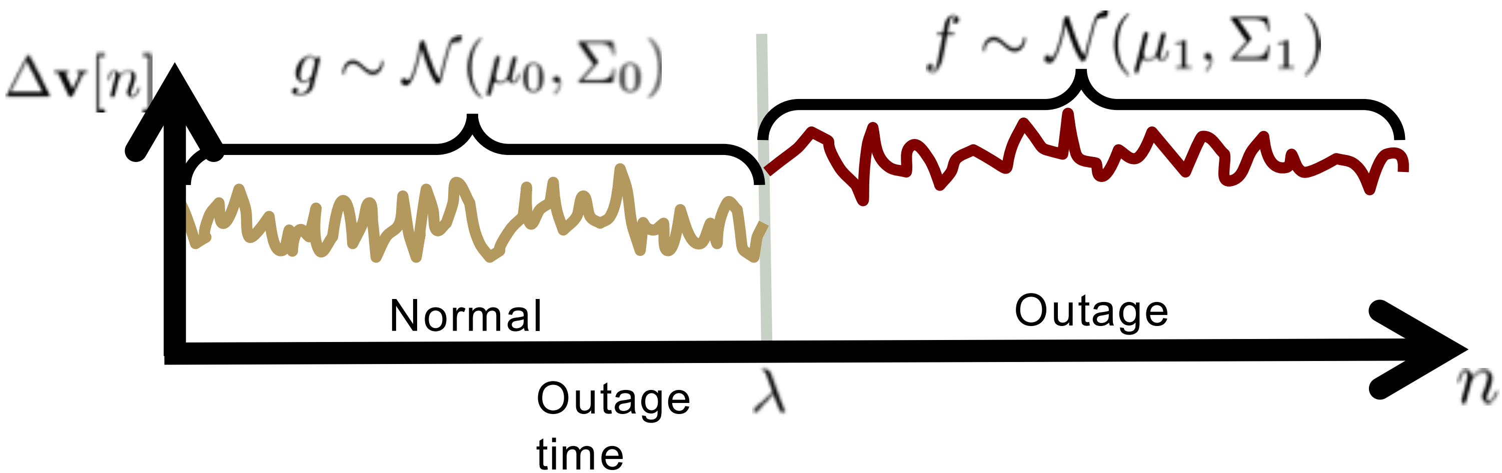


(1) Find: outage time λ

(2) Find: outage location

CIS/OMS Sensor Data Plug Validation (cont'd)

- Outage Detection: Probabilistic Description



Data likelihood ratio test

$$\Lambda(\Delta \mathbf{v}^{1:N}) = \frac{P(Normal|\Delta \mathbf{v}^{1:N})}{P(Outage|\Delta \mathbf{v}^{1:N})}$$

Normal: 0.1, 1, 10

Outage: 10^4, 10^5, 10^6

CIS/OMS Sensor Data Plug Validation (cont'd)

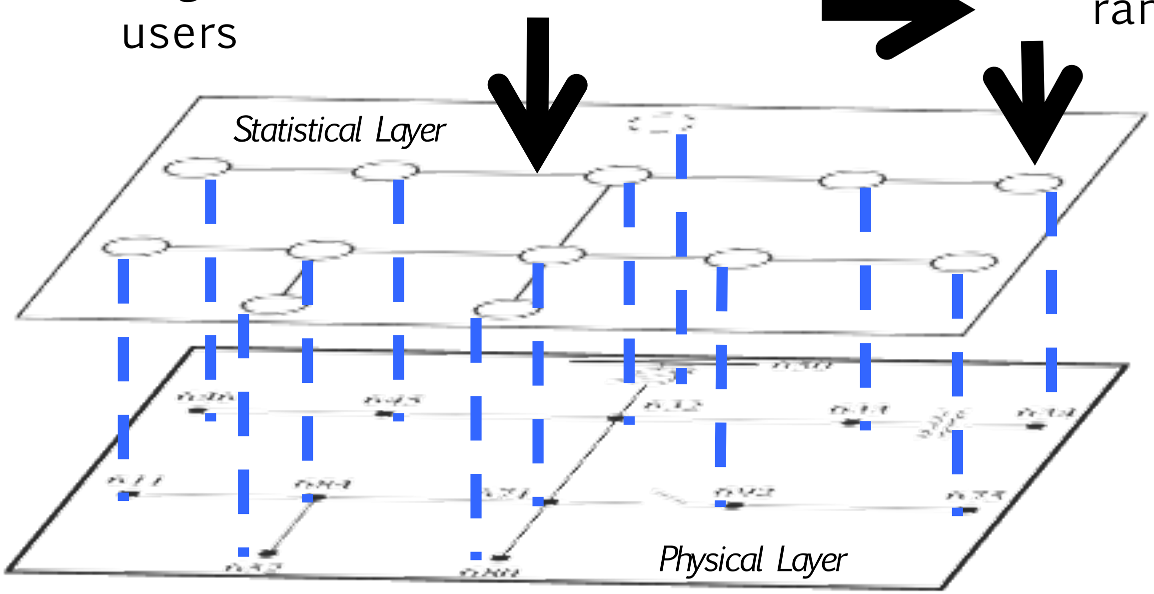
- Status of milestone: (M3.2.1) is slightly behind schedule; data access from utility will continue
- Path forward: API access to CIS / OMS data, verify the outage detection algorithm on both IEEE8-bus and 123-bus systems with and without DERs (also for Task 7.3)

GIS and feeder information (cont'd)

- Topology Identification: Statistical Learning

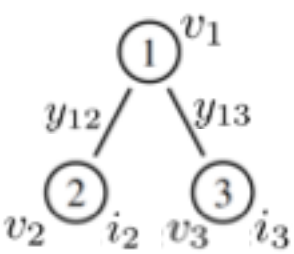
Edge: statistical link of users

Node: end-user data random variable (V_i)



GIS and feeder information (cont'd)

- Our Statistical Approach



$$v_2 = \frac{i_2}{y_{12}} + v_1 \quad v_3 = \frac{i_3}{y_{13}} + v_1$$

$$I_2 \perp I_3 \Rightarrow V_2|V_1 \perp V_3|V_1 \Rightarrow$$

$$P(\mathbf{V}) = P(V_1)P(V_2|V_1)P(V_3|V_1) = P_a(\mathbf{V})$$

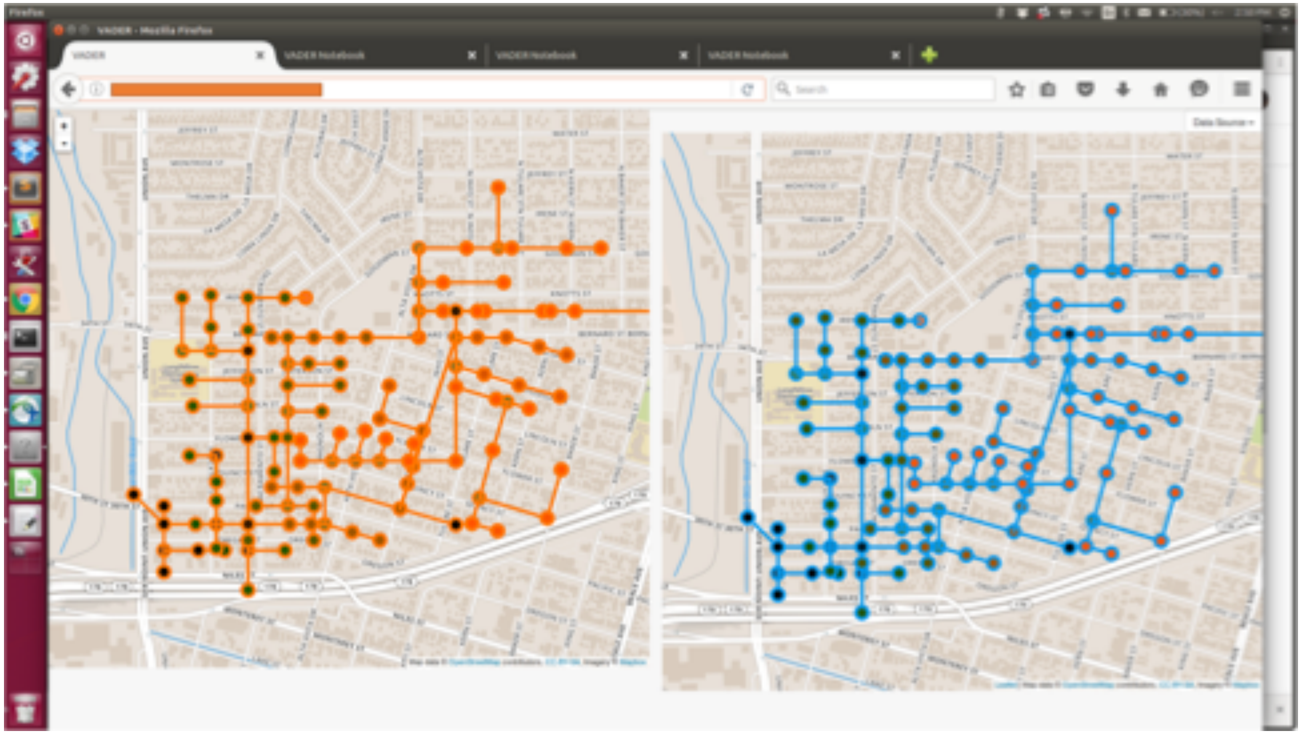
$$\left.\begin{array}{l}\text{tree network}\\ I_i \perp I_j, i \neq j\end{array}\right\} \Rightarrow V_i|V_k \perp V_j|V_k \Rightarrow$$

$$p(\mathbf{V}) \underset{\text{proof}}{\boxed{=}} p_a(\mathbf{V}) = \prod_{i=2}^{n} p(V_i|V_{\mathrm{pa(i)}})$$

GIS and feeder information (cont'd)

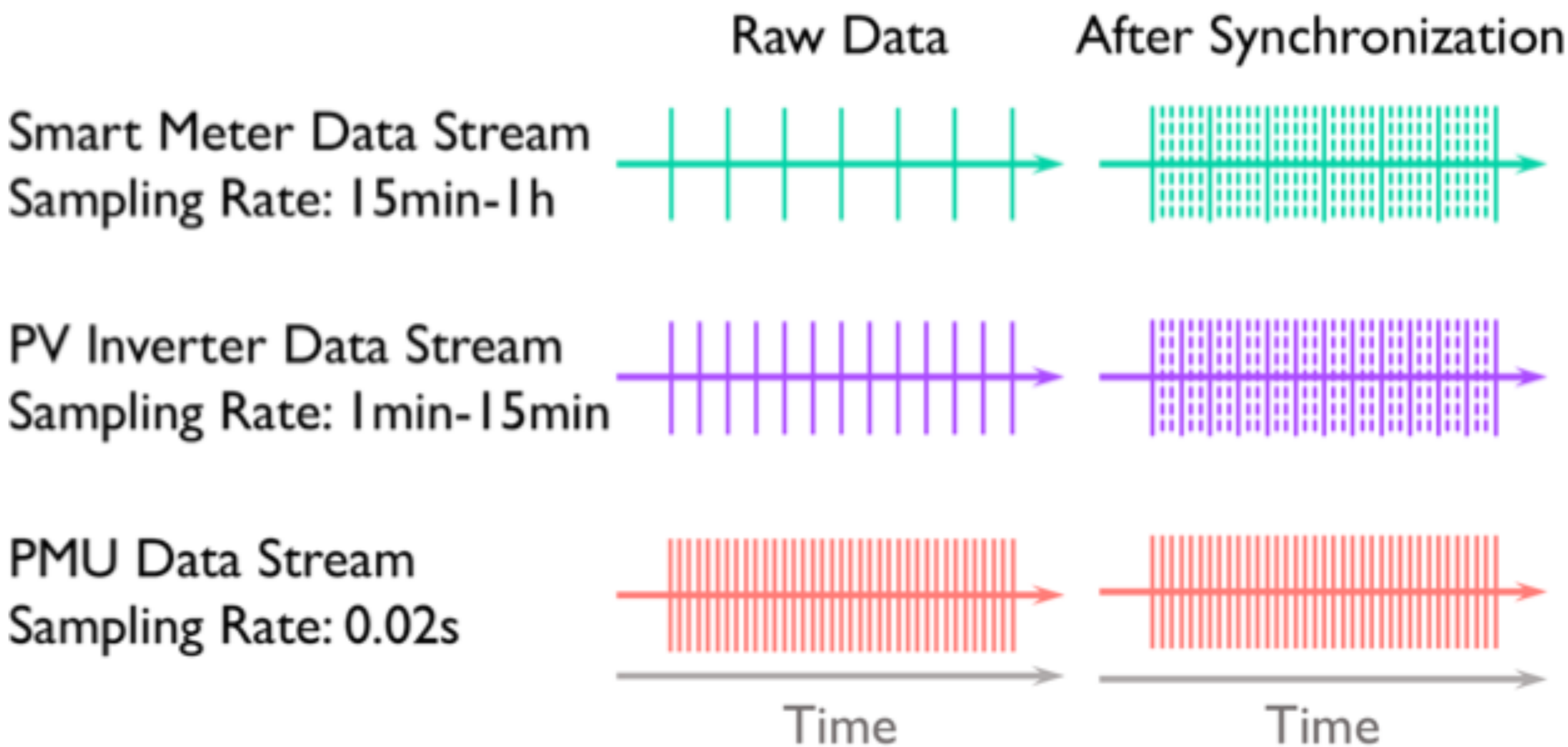
- Topology Identification Example

True Estimated



Virtual SCADA capability for data

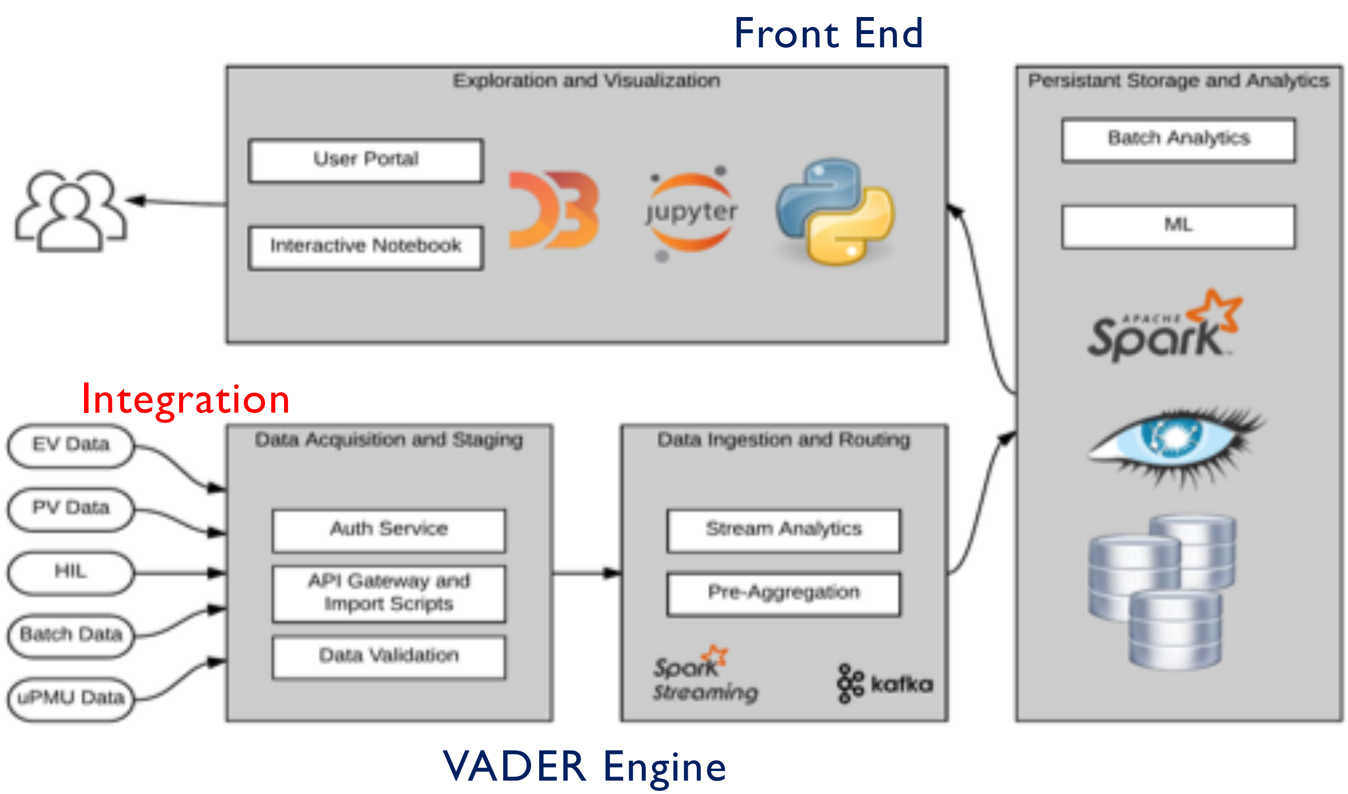
- Illustration of a "virtual-SCADA" use case (also for Task 3.1)



Task 4.1: System architecture (cont'd)

- Integration: a unified framework for planning & operations

- Revised *system architecture* in Q3: publish-subscribe messaging
  - A data pipeline ingesting data from the publishers and routing it to various subscribers



System architecture

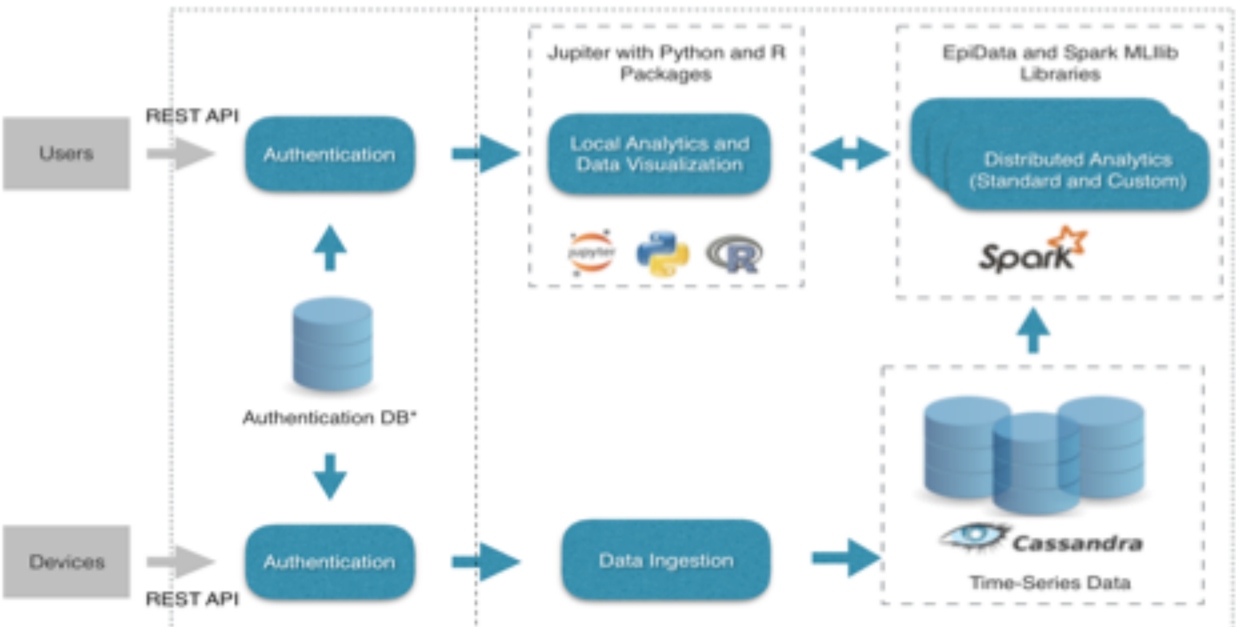
Data ingestion performance

	Run 1	Run 2	Run 3
Number of Records	1,000	2,000	10,000
Min Ingestion Time (ms)	23.25	22.20	21.13
Max Ingestion Time (ms)	257.18	296.36	1,199.94
Average Ingestion Time (ms)	30.47	30.46	31.30

- Persistent storage and analytics:
  - Data stored in Cassandra database; database schema architected for storing time-series data, spatial & geographical data
  - Apache Spark for computing batch analytics; integrate with Python & R through PySpark and SparkR for batch processing
- Exploration and visualization: Two user interfaces
  - User Portal; access to data using charts, tables
  - Interactive Notebook; custom queries and analytics

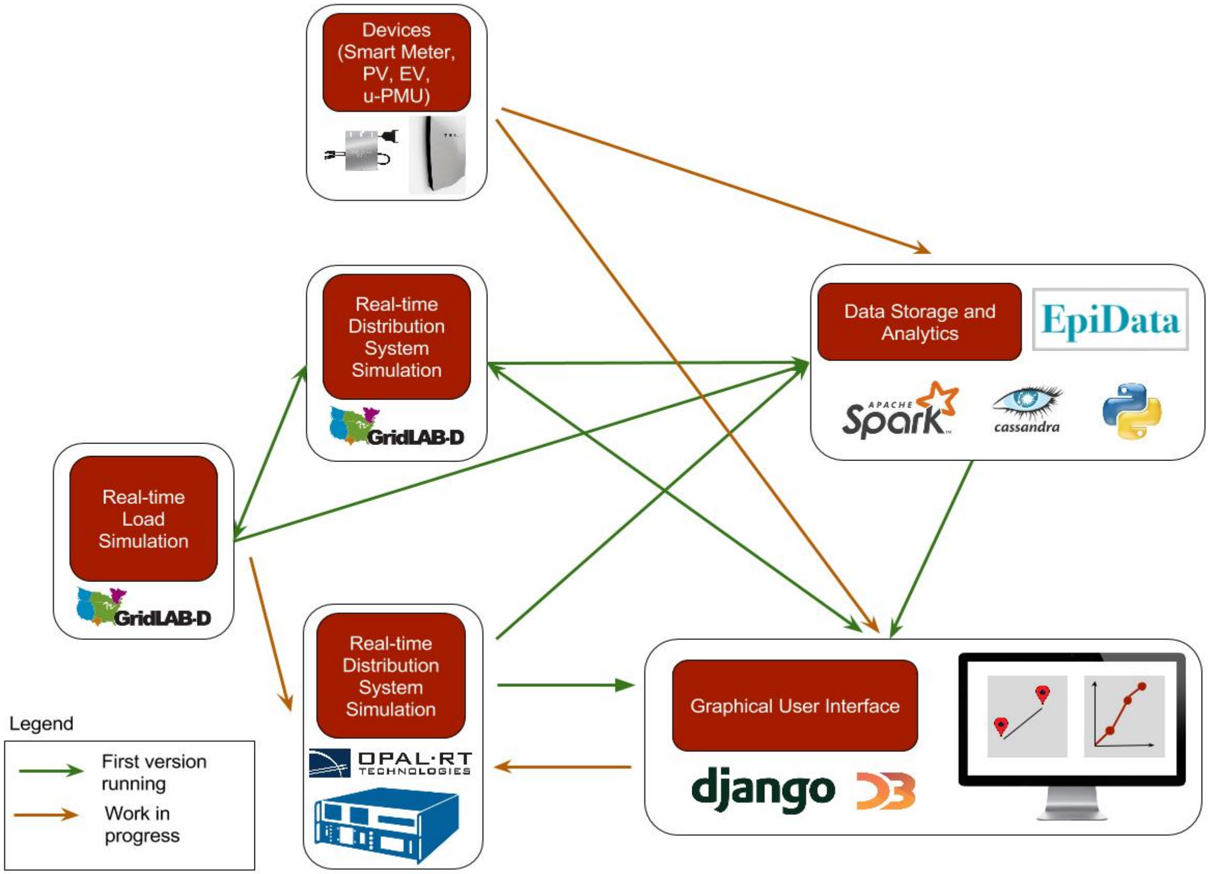
System architecture

- EpiData has developed an architecture for VADER *backend* that relies on open source "BigData" tools for information storage, retrieval, and processing. Open source tools:
  - Cassandra distributed database
  - Spark computing engine for ML analytics engine



VADER engine (cont'd)

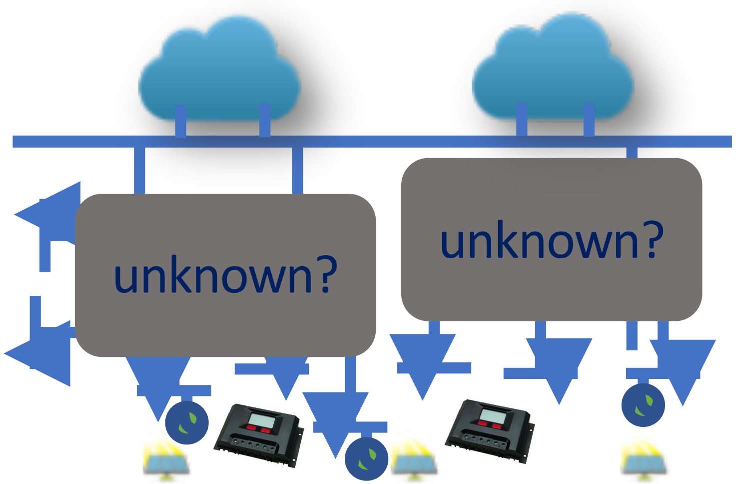
- *System integration*: VADER engine has been integrated with systems that provide input data

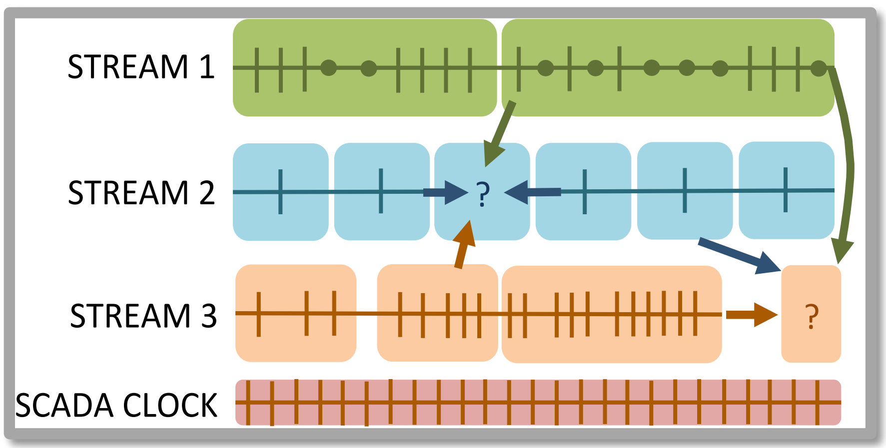


Integrate virtual SCADA with existing state estimation tool

- Traditional state estimation relies on knowing grid information; equations (power flow mapping) gives closed-form solution
- How to estimate states when there is lack of grid information, voltage violation, un-modelled active control?
- Data-driven approach mapping rules; Support Vector Regression

Mapping Rule: What now analytics

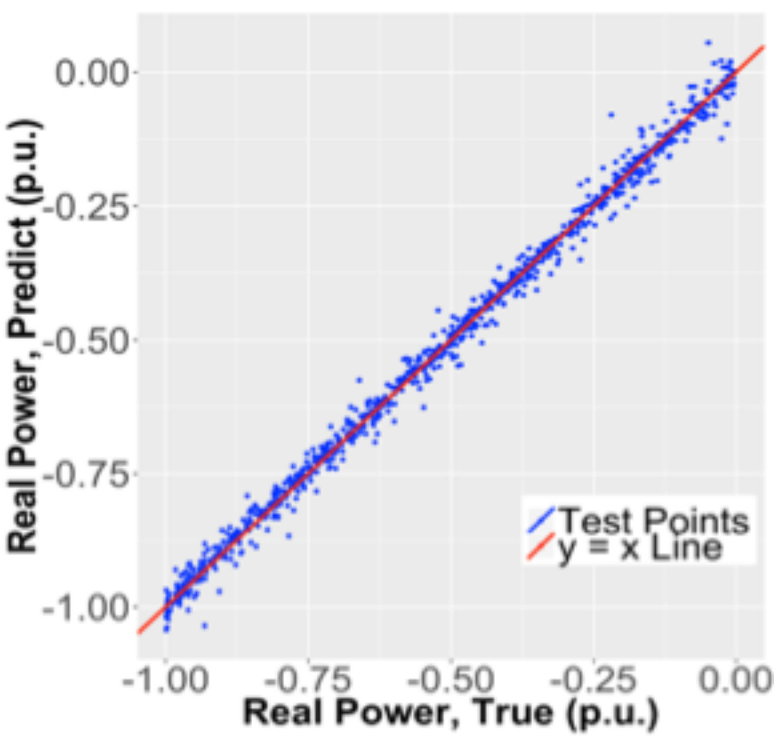




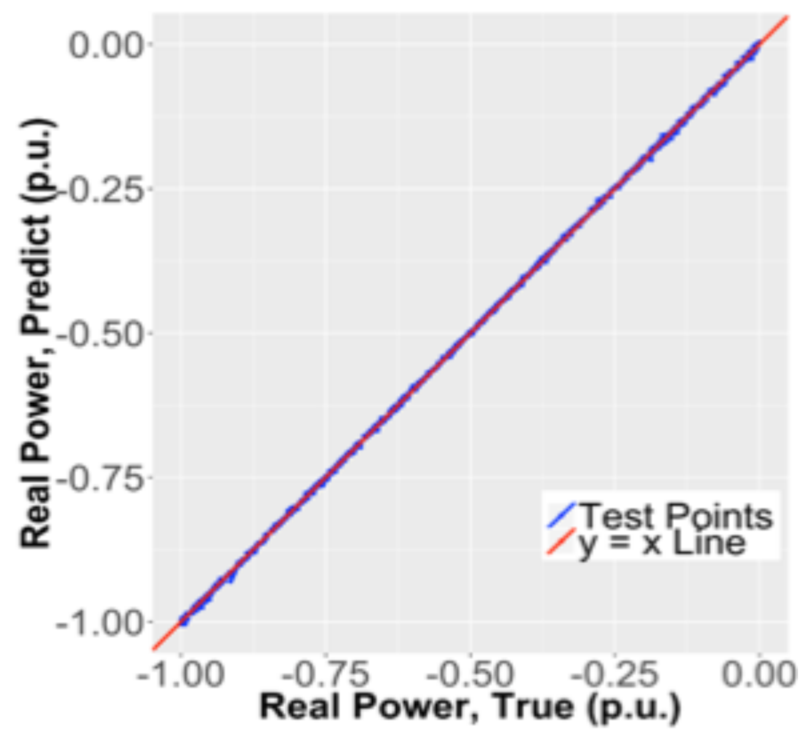
Integrate virtual SCADA with existing state estimation tool (cont'd)

- Simulation test results using IEEE 123-bus distribution grid: linear vs nonlinear models

Linear Model



Nonlinear model: SVR with 2nd poly kernel



Integrate virtual SCADA with existing state estimation tool (cont'd)

- Simulation test results: training and testing data ranges

Same range for training and testing data	Different Ranges	Different Ranges
Linear Model	Linear Model	SVR with 2^{nd} poly kernel