Data Science for Future Energy Systems

Sanmukh R. Kuppannagari
University of Southern California
Data Science Lab (**Prof Viktor Prasanna**) - dslab.usc.edu
kuppanna@usc.edu
https://sraok.space

Outline

- Future Energy Systems
- Data Science for Future Energy Systems
  - Data Science in Smart (Power) Grids
  - Data Science in Smart Oilfields
- Concluding Remarks

Outline

- Future Energy Systems
- Data Science for Future Energy Systems
  - Data Science in Smart (Power) Grids
  - Data Science in Smart Oilfields
- Concluding Remarks

Future Energy Systems (1)

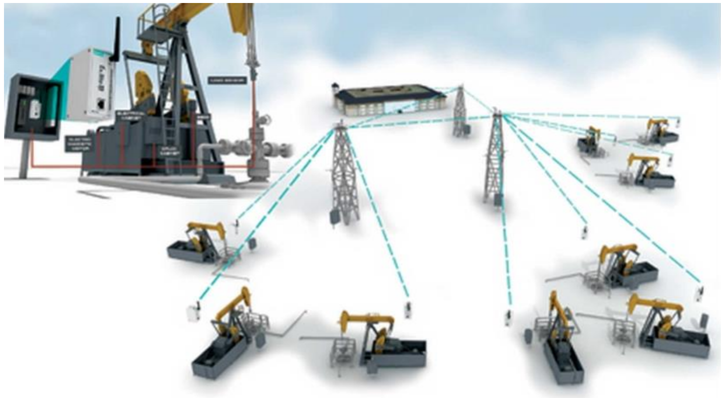
- Rapid Urbanization – increase in energy consumption
  - 6.5 Billions will live in urban areas by 2050
- Several Challenges still abound
  - 1 in 7 lack electricity access
  - 60% of green house gases comes from energy sector

Need novel, environmentally responsible solutions to the challenges resulting from rapid urbanization and population growth

Future Energy Systems (2)

Smart Grids

Smart Cities



Smart Oilfields

Future Energy Systems (3)

- Challenges
  - Unknown Complex System Dynamics
  - Need for Rapid Decision Making
  - High Cost of Failure

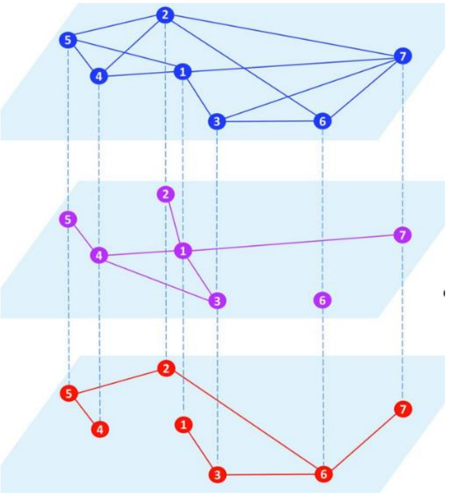
- Opportunities
  - Vast amount of available data from IoT enabled sensing and monitoring devices
  - Fast IoT enabled asset control especially with the upcoming 5G standard

- **Need for Data Science enabled Optimization for Future Energy Systems**

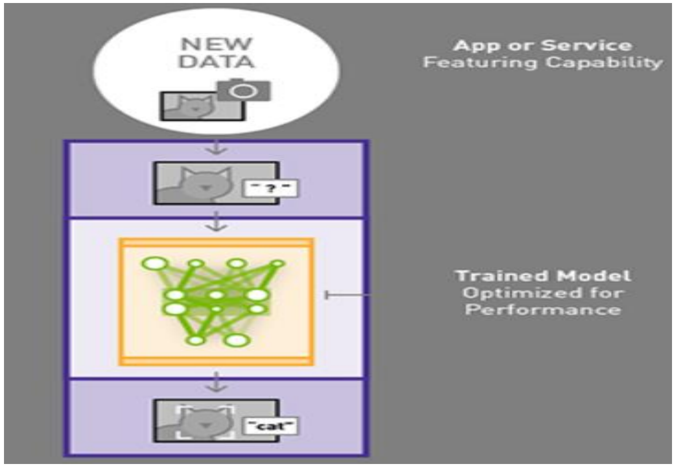
Outline

- Future Energy Systems
- Data Science for Future Energy Systems
  - Data Science in Smart (Power) Grids
  - Data Science in Smart Oilfields
- Concluding Remarks

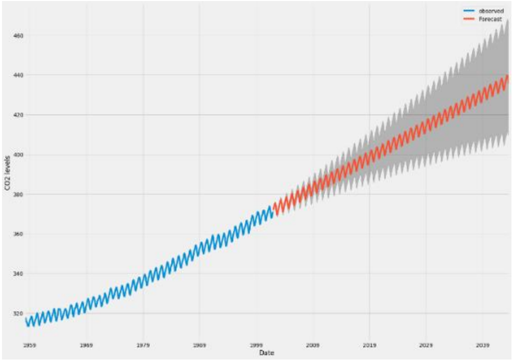
Data Science



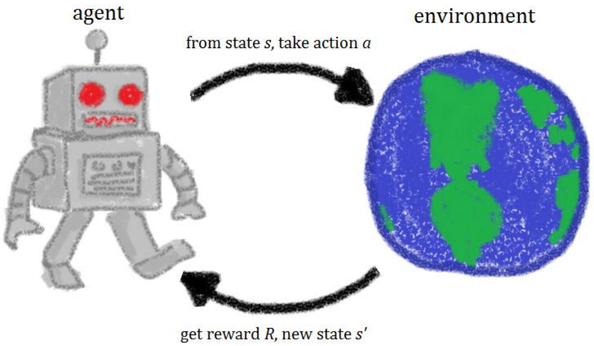
Modeling



Learning



Prediction



Optimization

Outline

- Future Energy Systems

- Data Science for Future Energy Systems
  - Data Science in Smart (Power) Grids
  - Data Science in Smart Oilfields

- Concluding Remarks

DEEP SOLAR

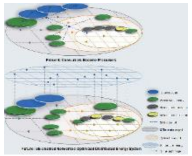
Data DrivEn Modeling and Analytics for Enhanced System Layer ImPlementation

http://deepsolar.usc.edu

Objective: Enable deep penetration of solar in distribution system

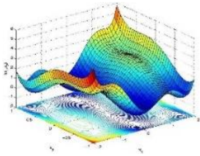
Challenges

- Partial observability of distribution system
- Stochasticity of solar generation
- Real time operational requirement



Data Driven Grid Modeling

Fast Robust Predictive Analytics



Realtime Scalable Optimization

What if Scenario Analysis

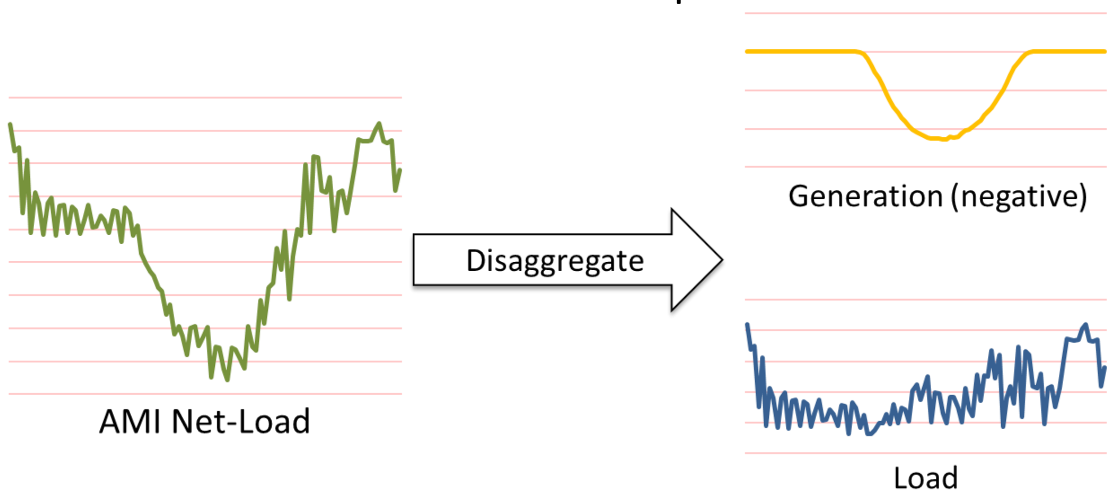
Powered by SunShot
U.S. Department of Energy

Behind-the-Meter Solar Generation Disaggregation (1)
Background

- **Problem:** Disaggregate *AMI Net-Load*

  AMI Net-load = Consumption - Generation
- **Inputs**
  - AMI data for each user
  - Solar Irradiance
  - Location information
- **Output**
  - Consumption
  - Generation

- **Challenges**
  - Ground truth unavailable -> unsupervised training
  - Many ways to disaggregate a given time series
  - Many latent factors that may affect consumption



Behind-the-Meter
Solar Generation Disaggregation (2)
Technology

Model

- AMI = load - generation
- Load modeling: linear combination of non-PV loads
- Generation modeling: linear function of solar irradiance

Approach

- Identity PV vs non-PV customers via K-means
- Clustering of non-PV customers to find centroids
- Solve *Disaggregation Optimization Problem*

Disaggregation Optimization Problem

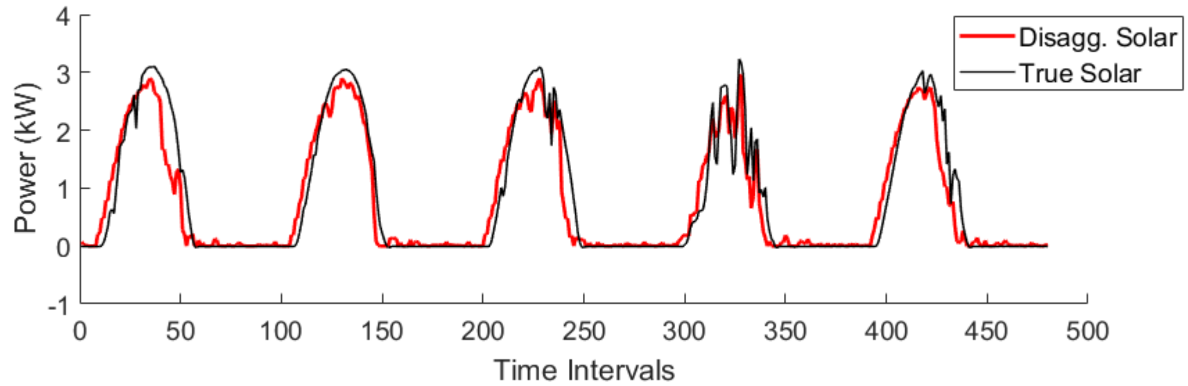
- Load $l(t) = \sum_i \theta_i g_i(t)$
  $g_i(t)$: centroids of clusters at time t
- Generation *g(t)* = $\phi R(t)$
  $R(t)$ solar radiance at time t

- Find θ, ϕ that minimizes:

$$\sum_{t=1}^{T} \left(\sum_{i} \theta_i g_i(t) + \phi R(t) - y(t) \right)^2$$

Behind-the-Meter
Solar Generation Disaggregation (3)
Results

- Pecan Street dataset: AMI data for 200 customers
- Error metric: Mean Absolute Scaled Error (MASE) (Normalized error)
- ~11% lower MASE for generation, ~21% lower MASE for consumption



C. M. Cheung, W. Zhong, C. X. Xiong, A. Srivastava, R. Kannan, V. Prasanna,
“Behind-the-Meter Solar Generation Disaggregation using Consumer Mixture Models,” IEEE Smartgridcomm 2018

Synthetic Data Generation (1) Background

- **Motivation:** Lack of large scale distribution network dataset due to
  - Rapid evolution in DER types and technology
  - Data availability from small scale testbeds only

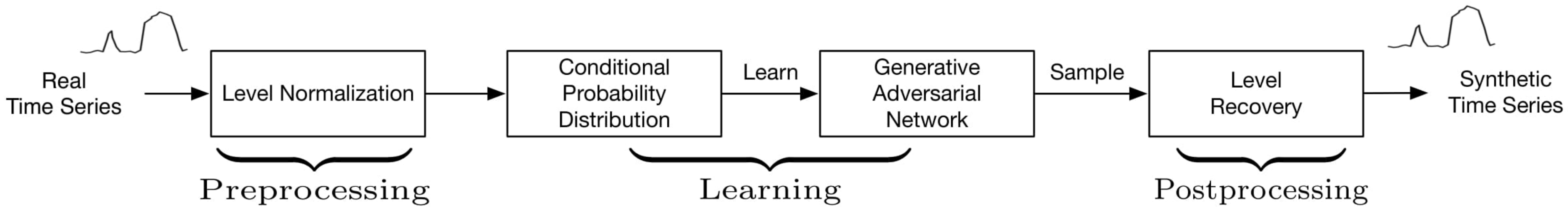
- **Problem:** Generate synthetic datasets for smart grid applications
  - **Input:** Small seed set of consumption and generation time series data
  - **Output:** Large set of synthetic consumption and generation time series data

- **Challenges**
  - Real life datasets contain too much noise
  - Data distribution gradually shifts over time

C. Zhang, S. Kuppannagari, R. Kannan, V. Prasanna, *Generative Adversarial Network for Synthetic Time Series Data Generation in Smart Grids*, SmartGridComm, 2018.

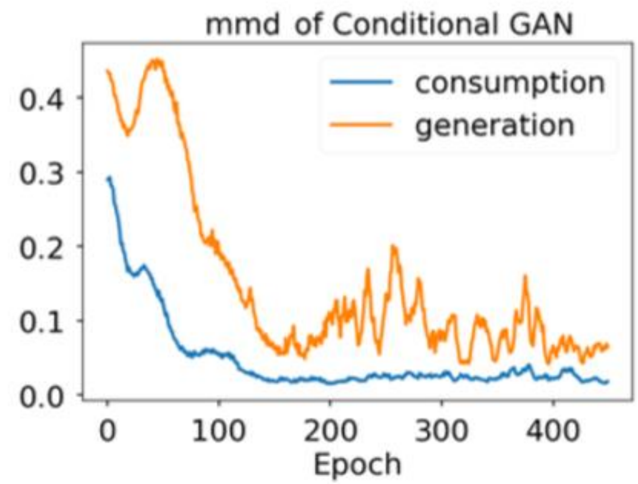
Synthetic Data Generation (2)
Technology

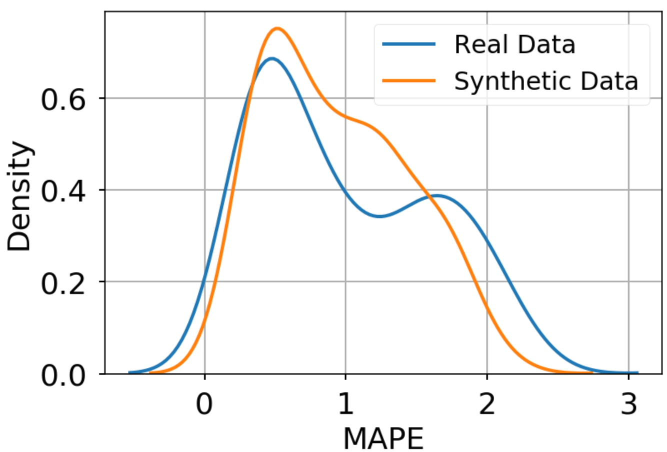
- **Modeling**
  - Timeseries = Level (mean) + pattern (noise)
- **Generative Adversarial Network (GAN):** Two contesting neural networks - sample generator vs evaluator
- **Approach**
  - Remove level from time series
  - Train Conditional GAN to learn pattern
  - Generate samples using GAN
  - Add mean and output to final time series



Synthetic Data Generation (3)
Results

- Dataset: Pecan Street
  - Load/generation data for ~200 customers
- Generate synthetic data for each real time series data
- Evaluation Metrics
  - Maximum Mean Discrepancy (MMD)
  - Real and synthetic dataset indistinguishable for ML tasks (clustering, prediction, etc.)





- **Results**
  - MMD converges towards small value -> real and synthetic from similar distribution
  - Similar Mean Average Percentage Error (MAPE) distribution for load prediction -> probability of incurring an error ϵ similar under real and synthetic data

Integrated Energy Management in Microgrids (1)

Background

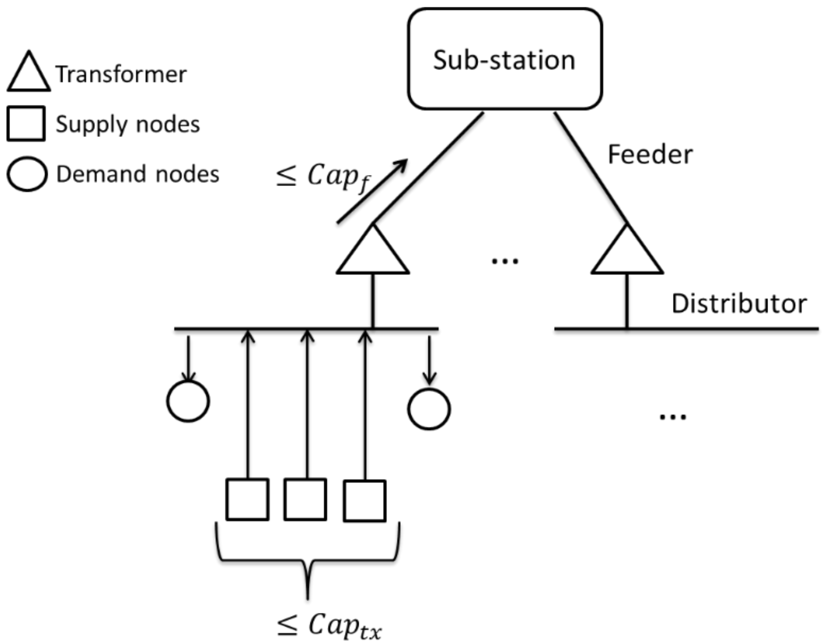
- **Motivation:** Single *Integrated* modeling and control of DERs in a microgrid
  - Distributed Energy Resources (DER): PV, EV, Storage, Smart Buildings
- **Problem:** Minimum cost DER scheduling in Microgrids
  - Cost: generation reward, curtailment cost, etc.
  - Scheduling: control the output of DER
  - **Inputs:** Cost, output, grid constraints

- **Challenges**
  - Large number of DERs and load
  - Computationally intensive
  - Constraints
    - Real time operational requirement
    - Voltage, power capacity

Integrated Energy Management in Microgrids (2)

Technology

- **Modeling**
  - DER output behavior learning and prediction
  - Physical constraints
    - Limit PV supply into distributors
    - Limit inverse power flow through feeders



- **Approach**
  - Reinforcement Learning + Dynamic Programming based approximation algorithm

Theorem: Integrated Energy Management with:

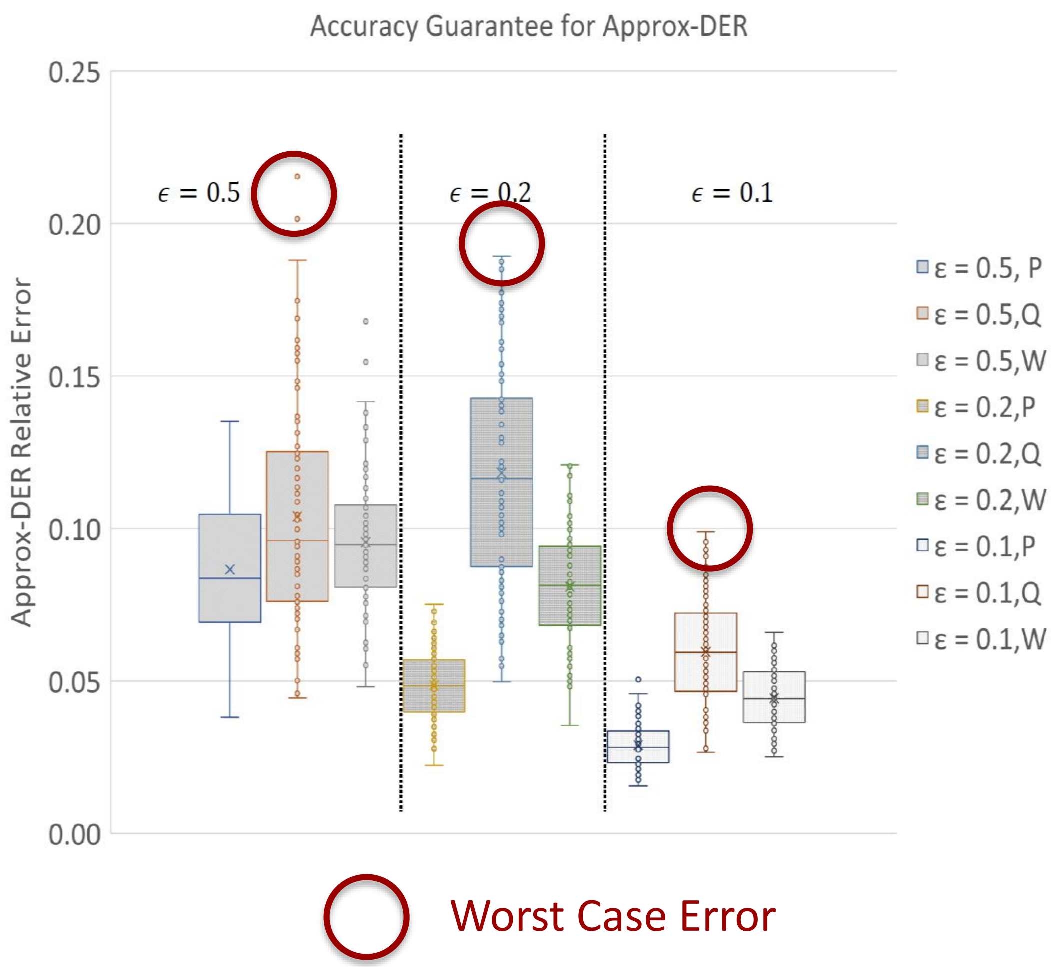
- $(1+\epsilon)$ bound on worst case error

Integrated Energy Management in Microgrids (3)

Results

- DER – PV systems
- Compare against optimal (ILP solution)

- Error less than theoretical bounds

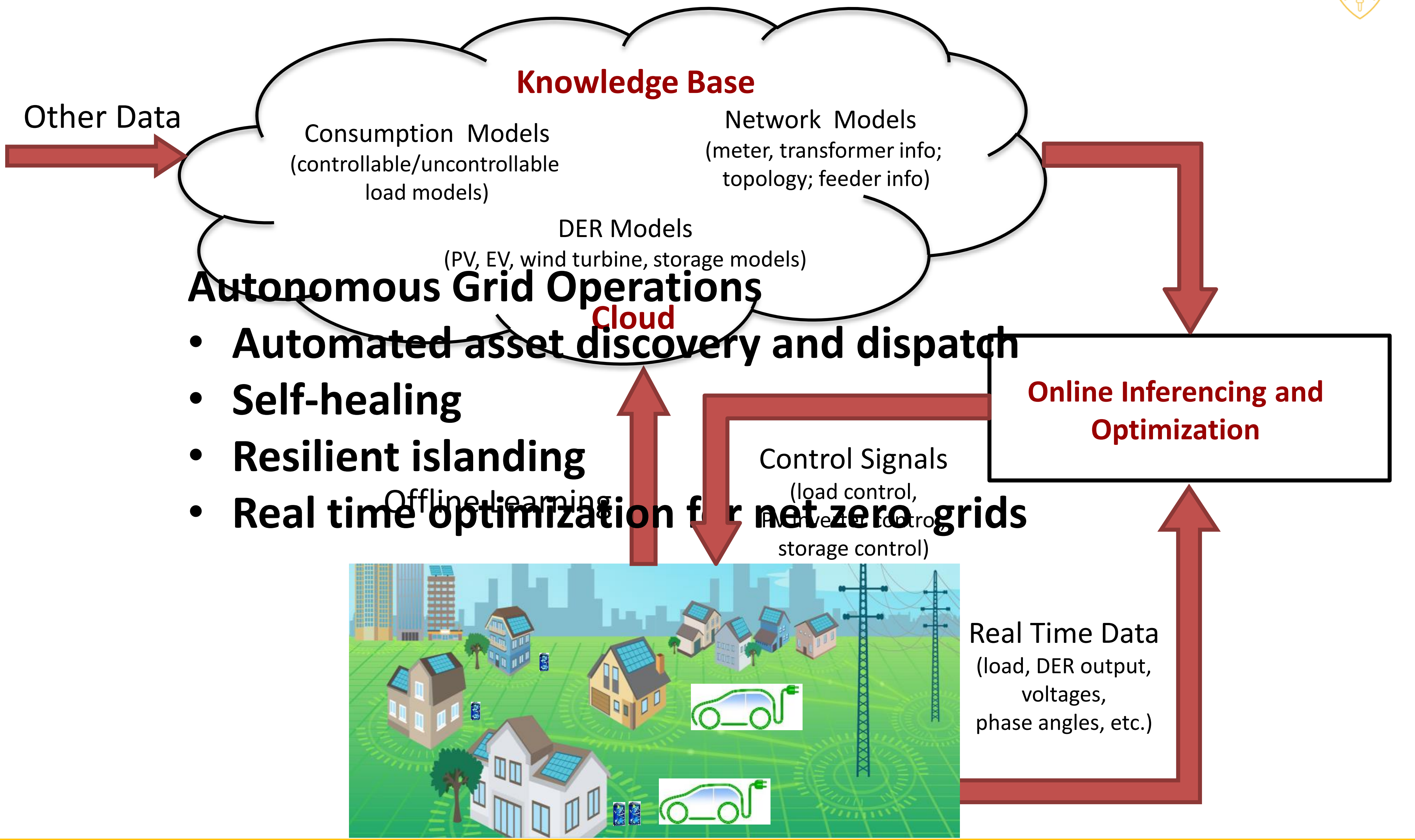


S. Kuppannagari, R. Kannan, and V. Prasanna Approximate Scheduling of DERs with Discrete Complex Injections, *e-Energy* 2019.

Towards Autonomous Energy Grids

Autonomous Grid Operations

- Automated asset discovery and dispatch
- Self-healing
- Resilient islanding
- Real time optimization for net zero grids



Outline

- Future Energy Systems
- Data Science for Future Energy Systems
  - Data Science in Smart (Power) Grids
  - Data Science in Smart Oilfields
- Concluding Remarks

Cyclic Steam Job Candidate Selection (1) Background

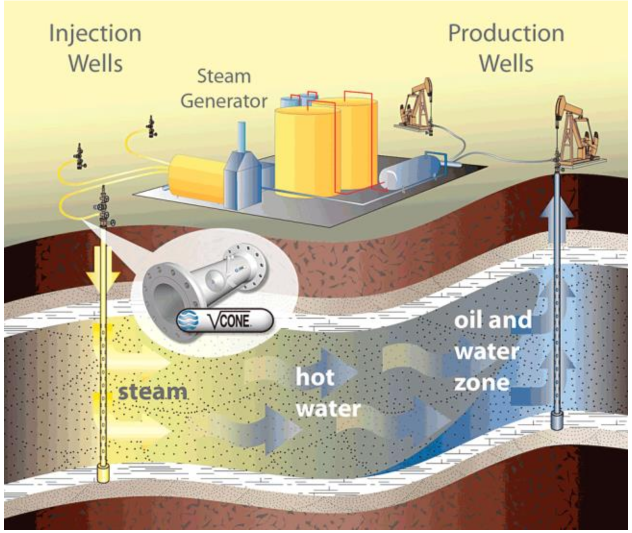
Motivation

- Injecting steam increases well temperature → Lowers oil viscosity → increases production

Problem

Predict production gain from a steam job

- **Input:** Time series data on production and well status before a steam job
- **Output:** Predict % gain in production after a steam job



Challenges

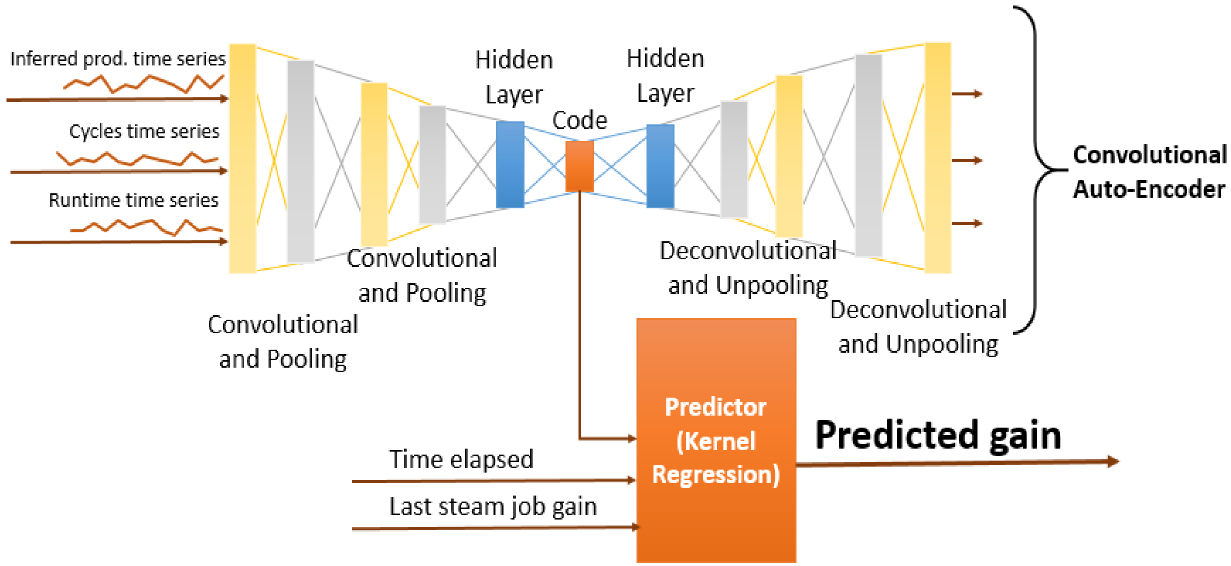
- High dimensionality of time series data
- Noisy data

Cheung, Chung Ming, Palash Goyal, Viktor K. Prasanna, and Arash Saber Tehrani. "OReONet: Deep convolutional network for oil reservoir optimization." In *Big Data (Big Data), 2017 IEEE International Conference on*, pp. 1277-1282. IEEE, 2017.

Cyclic Steam Job Candidate Selection (2) Technology

Approach

1. Encode timeseries with **Convolutional Autoencoder**
2. Input **autoencoder extracted** features with **manually extracted** features to a regressor
3. Regressor outputs predicted gain
4. Choose highest gain wells for steam job



Cyclic Steam Job Candidate Selection (3) Results

- Dataset
  - 4000 steam jobs in 1000 oil wells spanning 2014-2016

- Evaluation
  - Regression using
    - Baseline: Manually selected Features (MF)
    - Auto-encoder Features (AF)
  - precision@k: % of good predictions in the top k predictions

	Support Vector Regression		Kernel Regression	
	MF	AF	MF	AF
Mean Squared Error	15.4	2.61	10.4	0.59
Overlap coefficient (50)	0.44	0.68	0.36	0.74
Precision@50	0.6	0.98	0.56	0.98

Outline

- Future Energy Systems
- Data Science for Future Energy Systems
  - Data Science in Smart (Power) Grids
  - Data Science in Smart Oilfields
- Concluding Remarks

Concluding Remarks

- Data Driven Optimization Imperative due to Unknown System Dynamics of Energy Systems
- Novel applications of Data Science needed to address challenges of rapid urbanization and ensure sustainability
- Papers in this Workshop
  - On Using Graph Signal Processing for Electrical Load Disaggregation
  - Variations in Residential Electricity Demand across Income Categories in Urban Bangalore: Results from Primary Survey

Thanks

Data Science Lab (**Prof Viktor Prasanna**) - dslab.usc.edu

kuppanna@usc.edu

https://sraok.space