Prediction Models for Dynamic Decision Making in Smart Grids

Saima Aman

Committee

Prof. Viktor K. Prasanna

Prof. Cauligi Raghavendra

Prof. Cyrus Shahabi

USC
School of Engineering

University of Southern California

Saima Aman

Ph.D. candidate in Computer Science (2010-present)
University of Southern California (USC)
Research Interests: Data Science, Health Informatics, Energy Informatics

Master's in Computer Science
University of Ottawa, Ottawa, **Canada**

Bachelor's in Computer Engineering
Aligarh Muslim University, Aligarh, **India**

Dynamic Decision Making in Smart Grid

*dynamic means decisions are made a few minutes to a few hours before they are to be implemented

Smart Grid

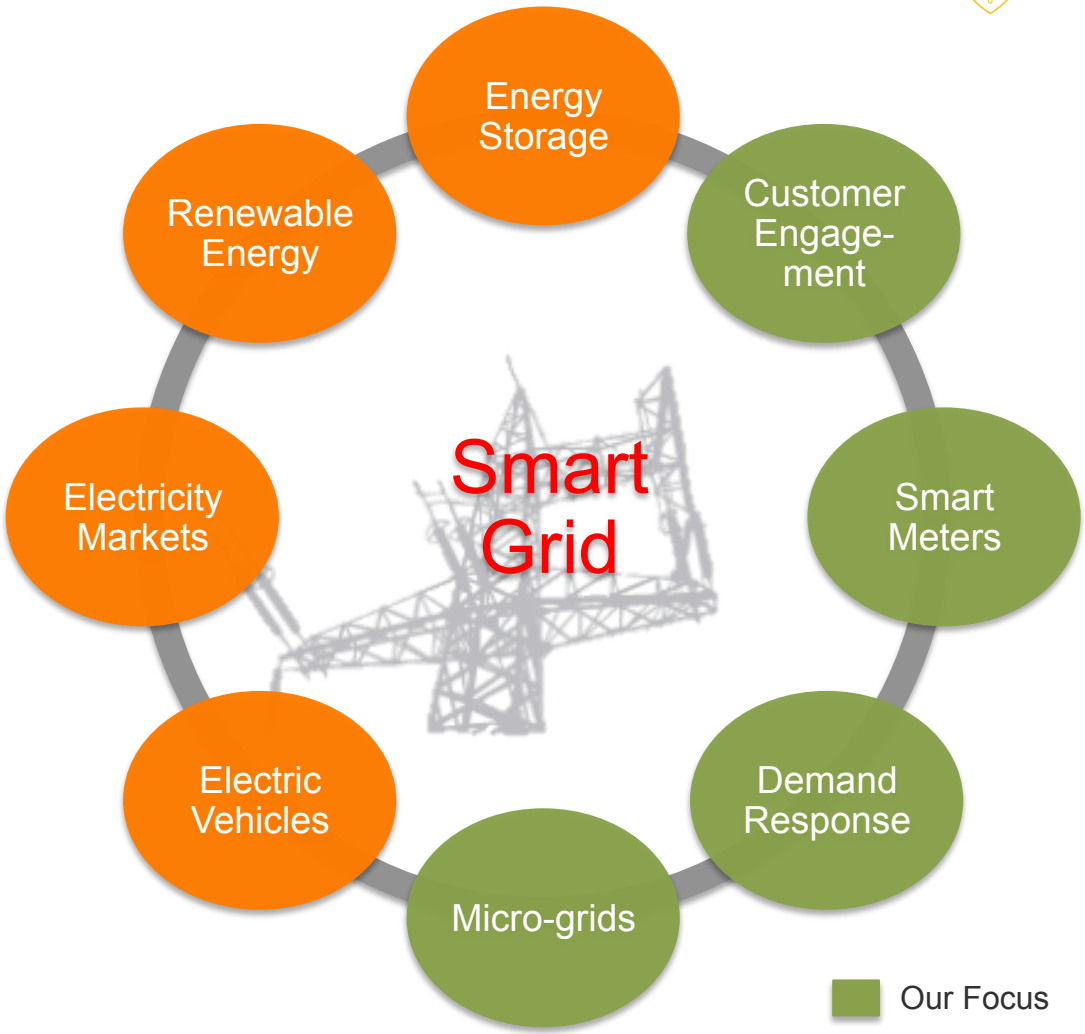
What?

Electric grid equipped with advanced technologies for

- monitoring
- control
- communication

Why

reliability
efficiency
sustainability



USC campus as a 'smart' microgrid

Motivation for our work

- Eliminate the need for manual intervention for demand optimization
- Enable automated decision making

Diversity

- Demographics
- Buildings (academic, admin, residential)

Scale

- 45K+ population
- ~50K sensors and smart meters
- 170 Buildings

Smart Equipment

- Measure energy usage at 1 min intervals
- Central control for zone temperatures and HVAC, VFD equipment, etc.

City within a City

Living Lab

Smart Grid Test-bed

Big Data Sources

Weather

24 readings per day

Electricity

170 meters in USC

50K in LA

96 readings in a day

Ambient Temperature

1000s sensors in USC

50K in LA

96 readings in a day

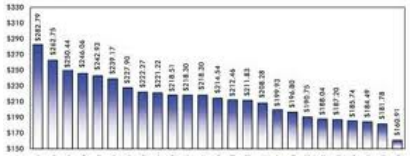
Sensors

Occupancy, light, thermal, etc.

50K sensors in USC

O(1mil) in LA

288 readings in a day



Physical features

170 buildings in USC

500K buildings in LA

Events

O(100) events per day in USC

O(10K) events per day in LA

Social media

50K people in USC

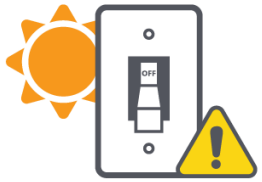
4 million people in LA

- Data is collected from sensors & other sources in real-time (every 15 minutes or less).
- Presents an opportunity to mine this data for actionable insights.

Demand Response (DR)

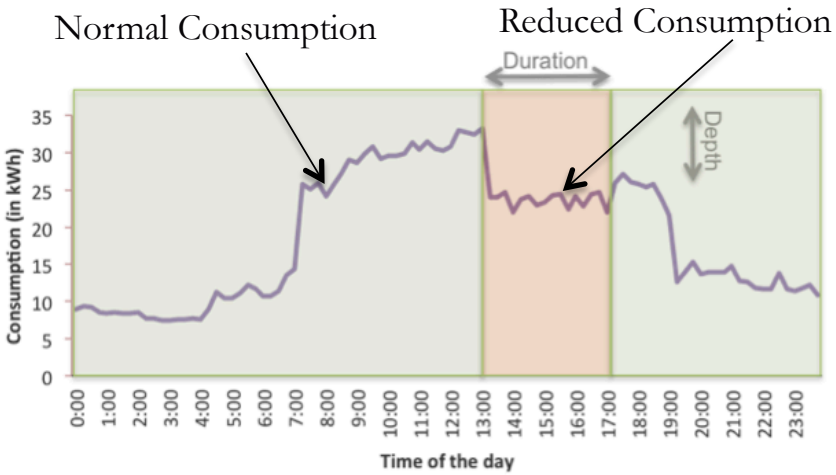
Peak Demand periods → Supply-demand mismatch → Service interruptions

Solution: Make the demand adaptive to supply conditions.



DR Event

- Utilities ask consumers to **decrease consumption** during anticipated peak demand periods.
- **Utilities** avoid the need to add additional generation units
- **Consumers**: get incentives in return



This works for 'anticipated' peak periods. Need to address "un-anticipated' peak periods.

Planning for DR

[day ahead] vs [hours/minutes ahead]

Planning for DR involves:

- Consumption prediction
- Decision making about when, by how much, and how to reduce consumption
- Sending notification to the customers

Day ahead planning

Traditionally, planning for DR is done **one day** ahead of the DR day. **(Ziekow et. al., 2013)**

Hours/Minutes ahead planning

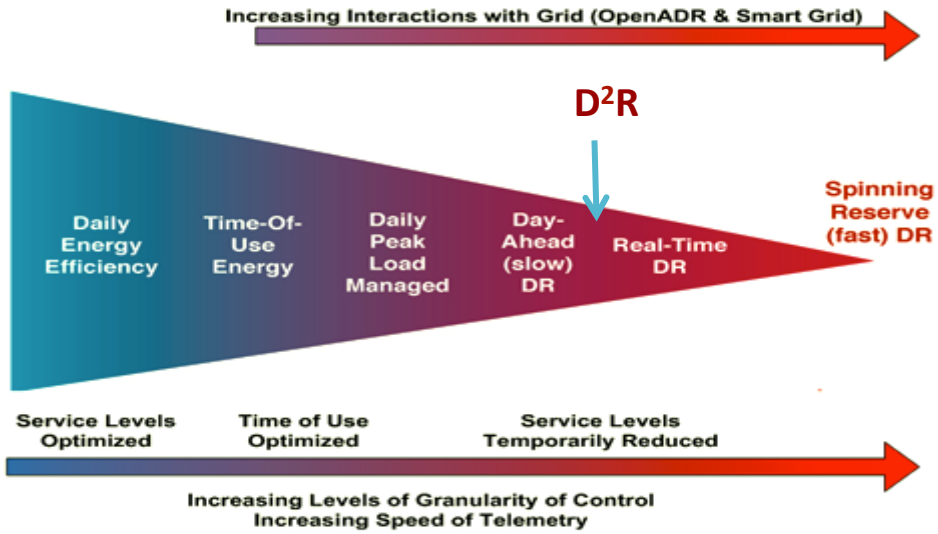
Needed due to dynamically changing conditions of the grid **(Simmhan et. al., 2013)**:

- Intermittent renewable energy sources
- Distributed energy sources
- Electric Vehicles
- Customer participation
- Special events

Factors driving the grid toward more dynamic operations

Proposing Dynamic Demand Response (D^2R)

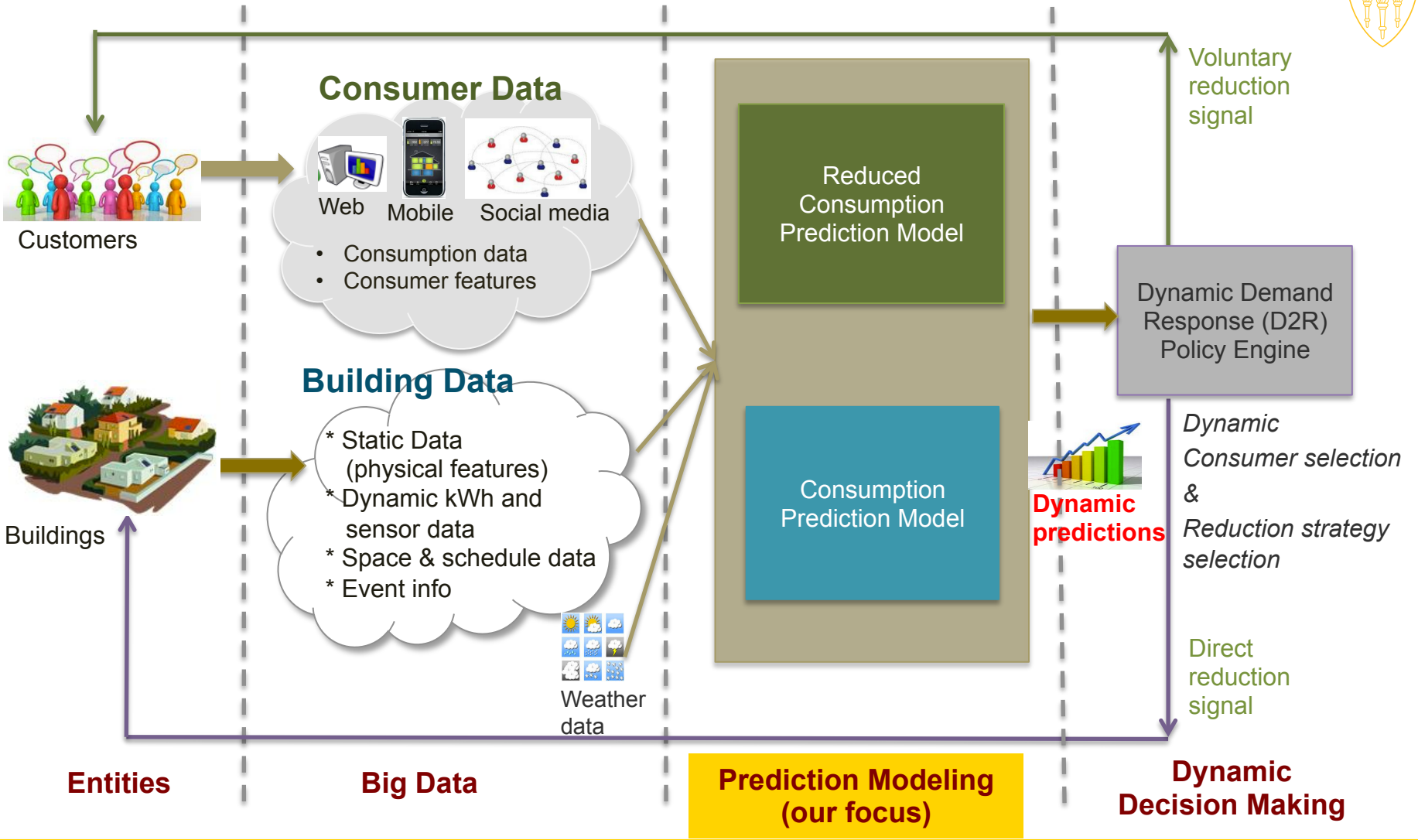
Dynamic demand response (D2R) is the process of balancing supply and demand in real-time and adapting to dynamically changing conditions by automating and transforming the demand response planning process. **(Aman et al., 2015)**



Source: Lawrence Berkeley National Lab

D^2R is a prime example of dynamic decision making in smart grid.

Prediction Models Help Enable D²R



Prediction Models for D²R Must Address Big Data Challenges

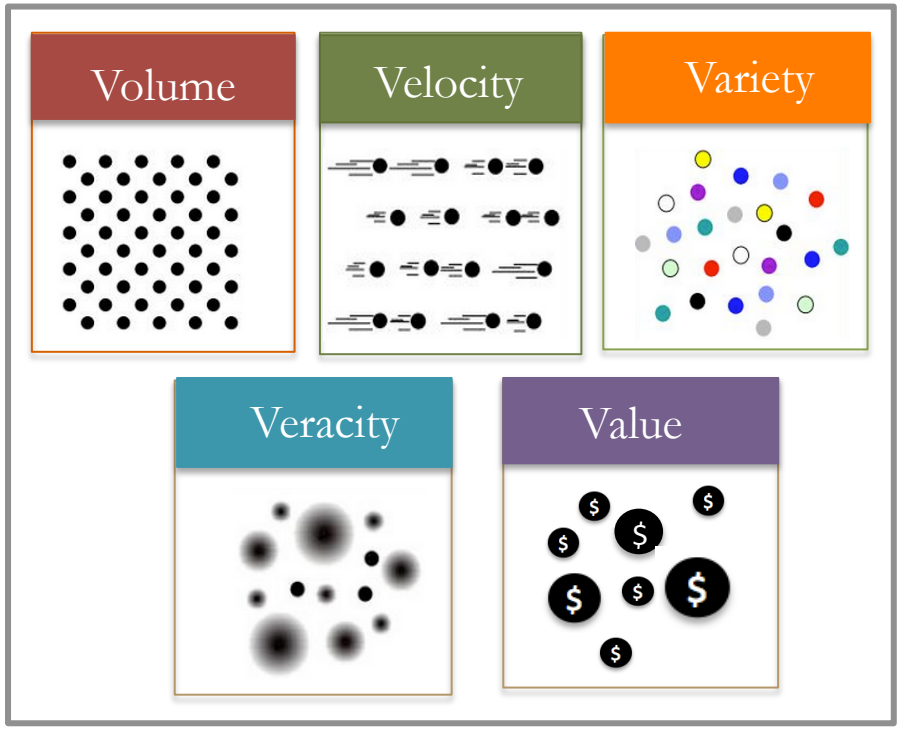
Feature Selection
- Relevant ones from large variety of features
- Parsimonious models preferred

Data Collection
- Effort required to acquire, assemble, and clean

Computational Complexity
- Time required in training and predictions is critical for dynamic predictions

Veracity
- Deal with imperfect data:
  - Missing data, partial data, etc.

Value
- Need to balance cost-benefit tradeoffs



5Vs of Big Data pose challenges for prediction.

Research Hypothesis

Prediction models utilizing big data can enhance dynamic decision making in smart grids.

Research Hypothesis

Prediction models utilizing big data can enhance dynamic decision making in smart grids.

Prediction models –

1) making predictions for the next few minutes to few hours horizon
2) evaluating prediction performance

Research Hypothesis

Prediction models utilizing big data can enhance dynamic decision making in smart grids.

big data –
Using data from a variety of sources and addressing the challenges of 5 Vs.

Research Hypothesis

Prediction models utilizing big data can enhance dynamic decision making in smart grids.

enhance –
our proposed prediction models help in some aspects of the decision making process, e.g., better accuracy with the available data, and faster decision making

Research Hypothesis

Prediction models utilizing big data can enhance dynamic decision making in smart grids.

decision making –
when, by how much, and how to reduce electricity use by the demand side

dynamic –
decisions are made from a few minutes to a few hours ahead

Research Contributions

Prediction with Partial Data

- Unavailability of data from sensors in real time leads to partial data
- We propose a novel model to predict for all sensors using only partial real time data from some 'influential' sensors

Prediction of Reduced Consumption

- Identify challenges of consumption prediction under DR
- We propose a novel ensemble that models "mean behavior" and "context dependent behavior" to predict reduced consumption during DR

Prediction Evaluation Measures

- Identify limitations of existing measures
- Propose a suite of evaluation measures addressing the following:
  - Dimension, Prediction bias, Scale, Reliability, Cost, Application-relevance

Contribution 1

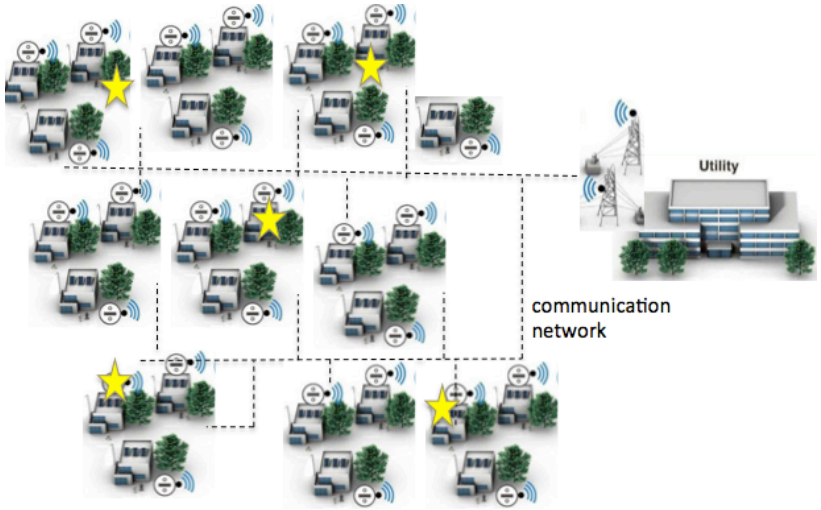
Prediction using Partial Data

Partial Data Problem

- Smart meters collect data in real-time (every 15 mins or less)
- Data is **not transmitted in real-time** to the utility, due to:
  - physical limitations of the transmission network (limited bandwidth)
  - security and privacy concerns of the consumers

- Only data from **some** meters (shown starred) is transmitted in real time.

- Complete data from **all** meters is available periodically when batch transmission takes place.

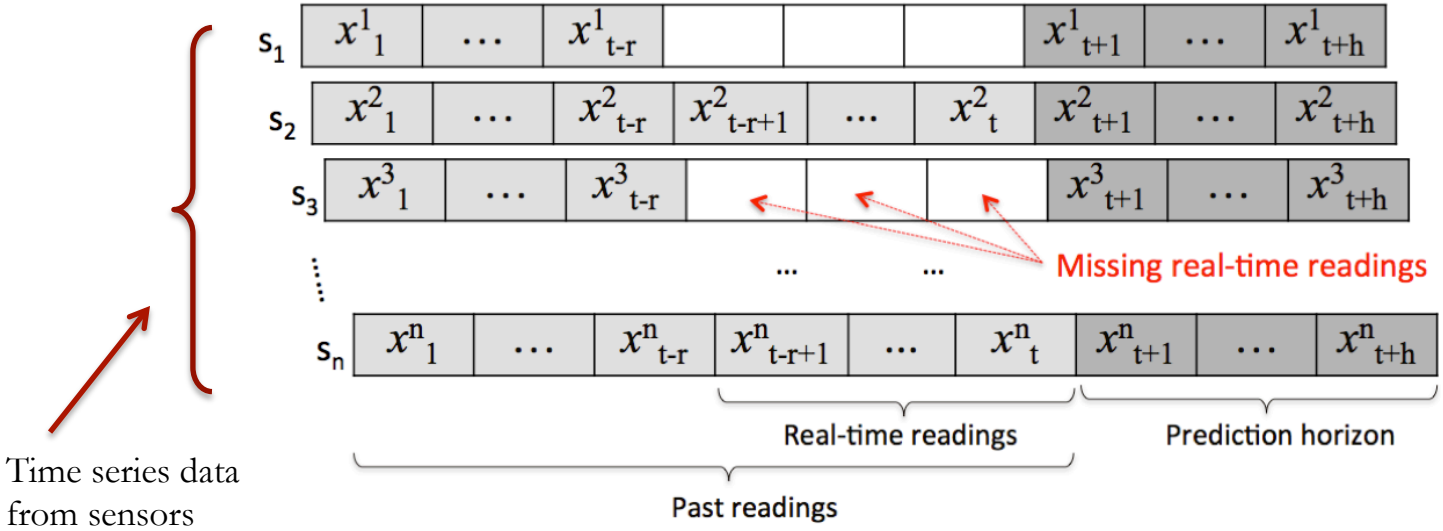


Only partial data is available in real-time.

Partial Data - Implications

Most prediction models are designed for ideal cases where all required data is readily available.

- Time-series models (e.g. ARIMA) and auto-regressive tree (ART) use recent real-time data Without real-time data, the performance of these models deteriorates.



- For dynamic demand-response, real-time data is critical to predict peak demands

Partial Data Vs Missing Data

	Missing Data	Partial Data
Timing	Unavailability of data at **arbitrary** time periods	Systematic unavailability of data for **known** time periods
Source	From **unknown** number of sensors	From a **known** subset of sensors
Cause	Due to **diverse factors**, such as faults	Due to **non-transmission of data** in that period
Recovery	Missing data is **lost**	Partial data becomes **available** when batch transmission occurs, and can be used to re-train our models
Related work	Missing data is estimated by interpolation methods **(Kreindler et. al., 2006), (Cuevas-Tello et al., 2010)**	None for partial data The volume of transmitted data is reduced by data compression **(Marascu, 2013)** or data aggregation **(Karimi et. al., 2013)**

Our approach – discovering 'influential' sensors

Instead of estimating unavailable real time data, we first discover influential sensors and use real time data only from them to do predictions for all sensors

We leverage the following:

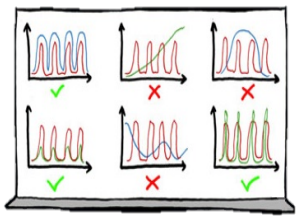
- Fine grained data logged locally at sensors – available periodically at the utilities
- Real-time data – always available from *some* sensors

} Used to select influential sensors

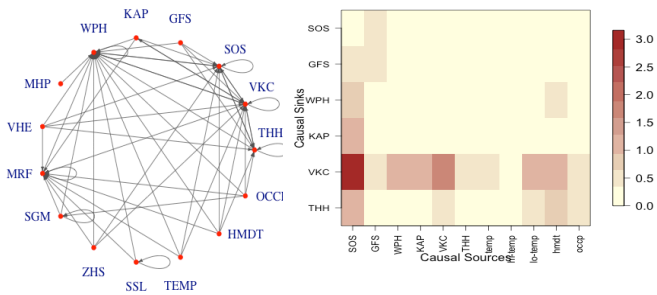
Hypothesis

Time series data of electricity consumption (and other schedule-driven data) shows dependencies

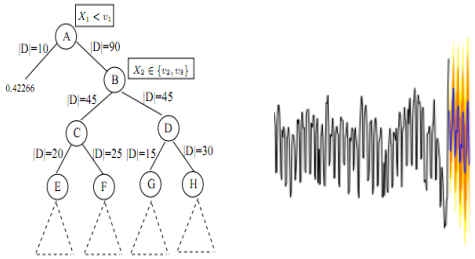
Our approach - Influence Model



Identify dependencies/ influence between time series from recent historical data



Identify sensors that show a stronger 'influence' on other sensors using the Lasso Granger method



Train regression tree models using real-time data from influential sensors as features

We use Lasso-Granger as a novel way of feature selection for regression tree. **(Arnold et. al., 2007)**

We use Lasso Granger for Influence Discovery

- Given n sensor outputs in form of time series $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n$
- Each series has observations at timestamps $\mathbf{t} = 1, \ldots T$

- For each series $\mathbf{x}^i$, a sparse solution for coefficients $\mathbf{w}$ is obtained by minimizing the sum of squared error and a constant times the L1-norm of the coefficients:

$$\mathbf{w} = \arg\min \sum_{t=l+1}^{T} \left\| x_t^i - \sum_{j=1}^{n} \mathbf{w}_{i,j}^T \mathcal{P}_t^j \right\|_2^2 + \lambda \|\mathbf{w}\|_1$$

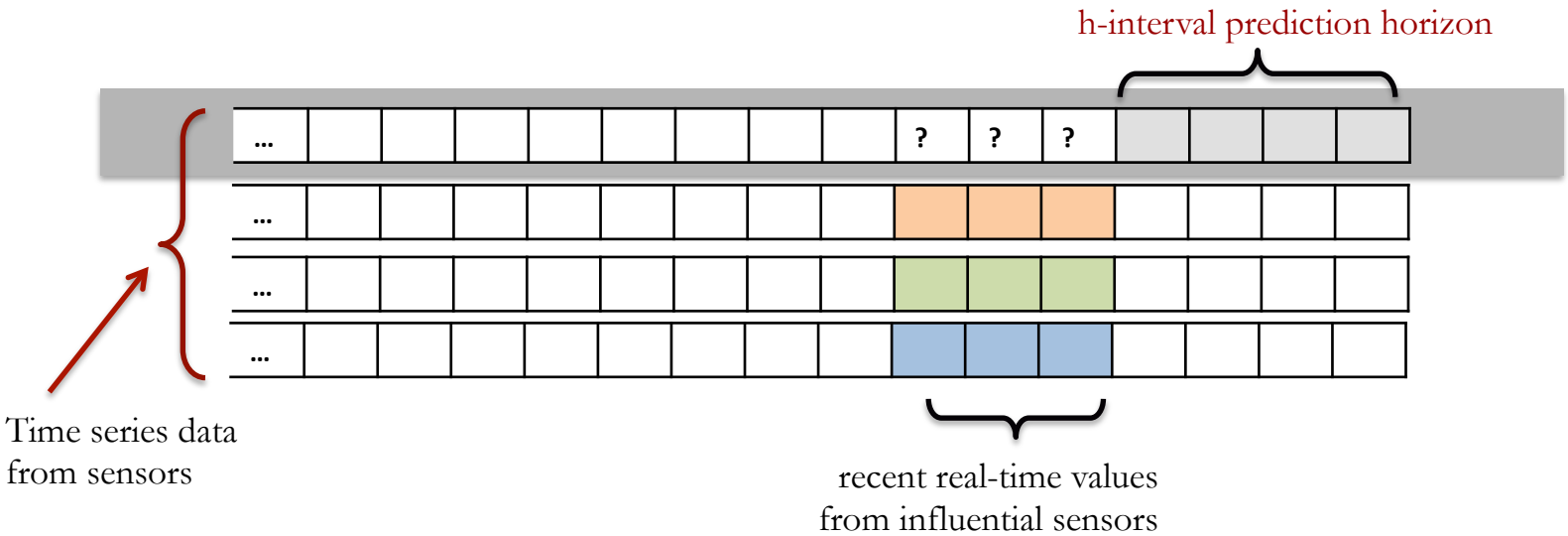
where $\mathcal{P}_t^j = [x_{t-l}^j, \ldots, x_{t-1}^j]$ is the sequence of past l readings

$\mathbf{w}_{i,j}$ is the coefficient representing the dependency of series i on series j
λ is the parameter that determines the sparseness of the coefficient vector $\mathbf{w}_i$

Lasso allows an efficient method for variable selection in high dimension **(Tibshrani '96), (Arnold et. al., 2007)**

Influence Model (IM)

- We propose Influence Model to solve the partial data problem **(Aman et. al., 2015)**
- Recent real-time values of other sensors are useful as predictors even in absence of the sensor's own real-time values
- A sensor's own relatively older values are less useful as predictors

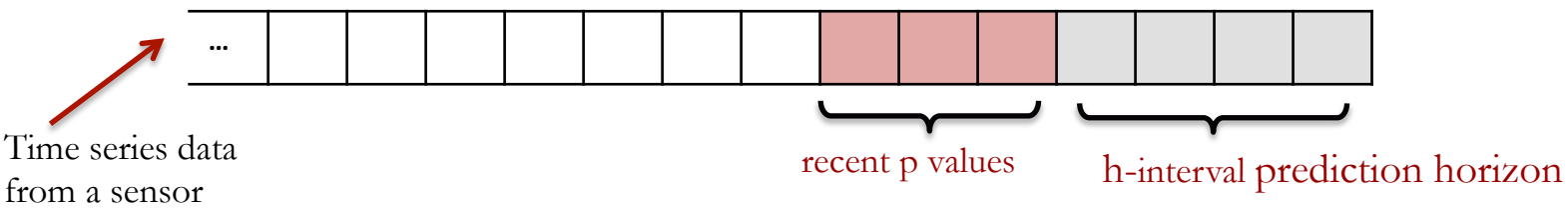


Baseline Models

Autoregressive Tree (ART)

- ART(p, h) – uses recent p values of a variable as features in a regression tree for h interval ahead prediction **(Meek et. al., 2002)**

- ART is a natural choice for baseline as it is also based on regression trees (like IM model)
- ART has been shown to offer high accuracy on large number of datasets **(Meek et. al., 2002)**



Local Influence & Global Influence Models

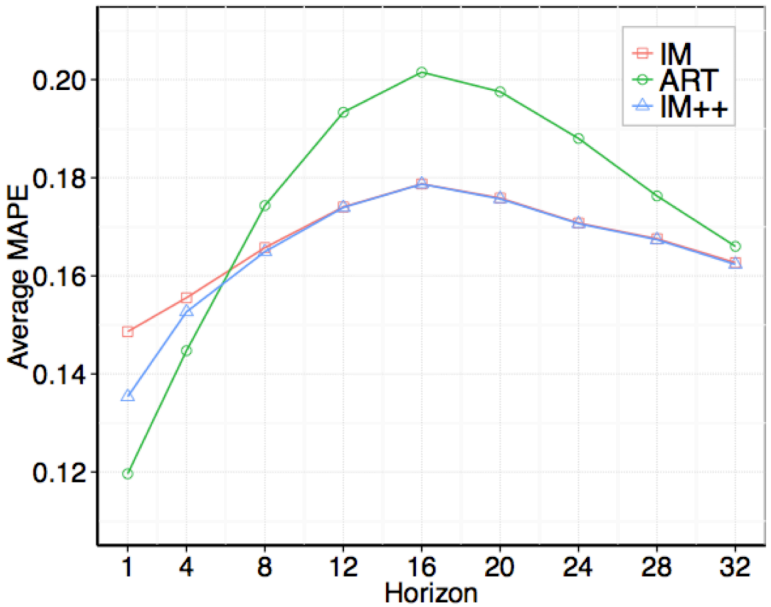
LIM (Local Influence Model)

- Without restricting the number of influential sensors, the selected influential sensors may include the total number of sensors.
- In LIM, we ensure that only a fraction of sensors is selected (top τ_l sensors selected locally).

GIM (Global Influence Model)

- Because local influencers are selected for each sensor, overall it may still result in a large number of sensors being selected
- In GIM, we select the top τ_g sensors globally.

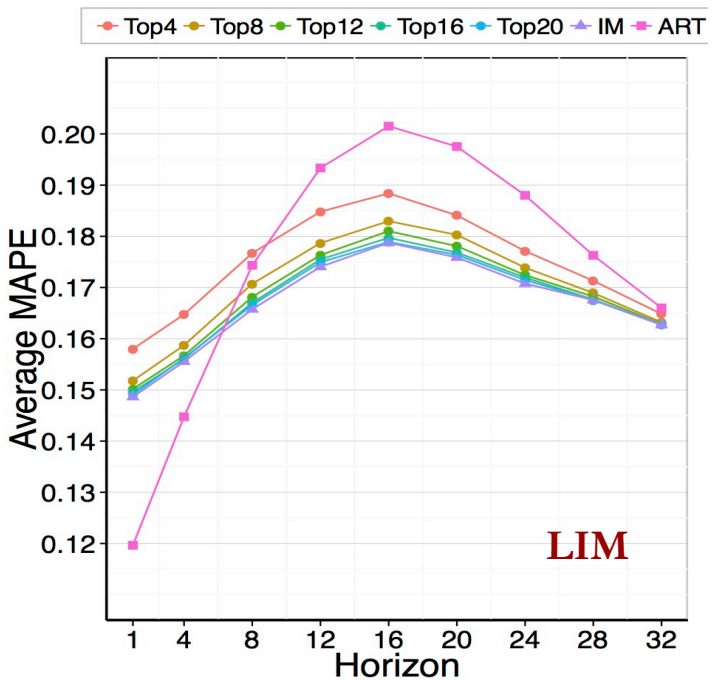
Results for IM

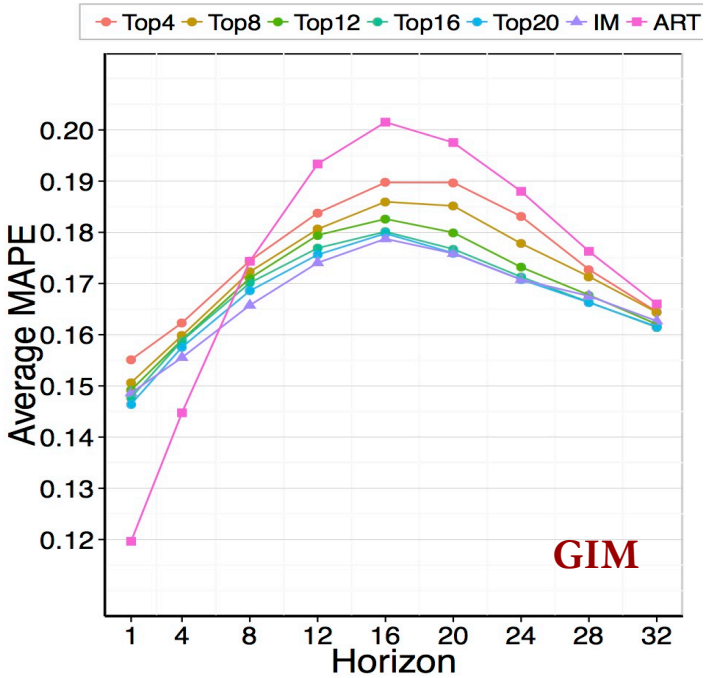


- 115 USC buildings
- 3 years' data
- @15-min intervals
- 8 hour prediction horizon

- Baseline (ART) performs well up to 6 intervals benefiting from real-time data.
- IM achieves comparable accuracy despite the lack of real-time data.
- IM's errors increase at a lower rate compared to ART
- With time, a sensor's own data becomes stale, and more recent real-time values of other sensors become more useful predictors.

Results for LIM and GIM





LIM (vs IM)

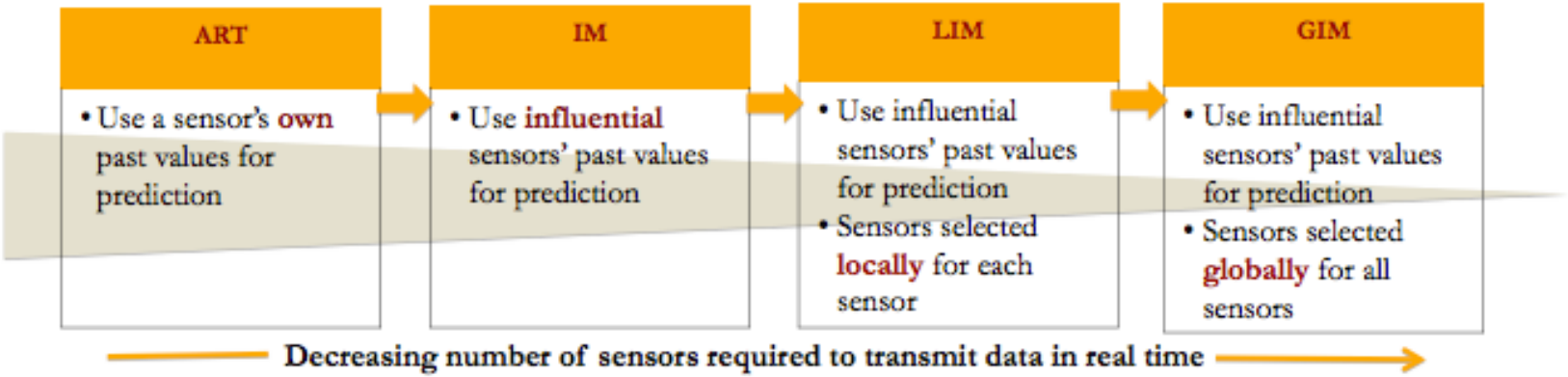
1.97% average increase for Top 8 and less than 1% increase for Top 12, 16, and 20 models.

GIM (vs ART)

Uses significantly lower number of sensors

Only ~0.5% increase in average error while using data from ~7% of sensors

Advantage of Influence Models



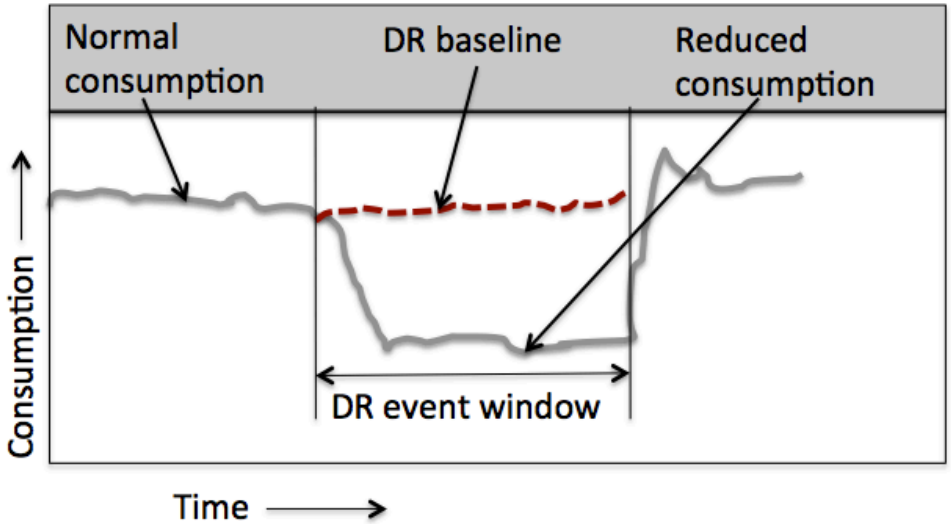
- ART requires real-time data from all sensors
- IM requires real-time data from only the influential sensors
- LIM requires real-time data from influential sensors selected locally for each sensor
- GIM uses real-time data from all influential sensors selected globally

Influential Models solve partial data problem in an efficient way without sacrificing accuracy.

Contribution 2

Prediction of Reduced Consumption

Prediction of Reduced Consumption



Reduced consumption prediction is useful in following **decision-making** tasks:

- estimating potential reduction during DR **(Chelmis et. al., 2015)**
- intelligently targeting customers for participation in DR **(Ziekow et. al., 2015)**
- performing dynamic DR at a few hours' notice **(Aman et. al., 2015)**
- estimating the amount of incentives to be given to the customers **(Wijaya et. al., 2014)**

Characteristics and Challenges

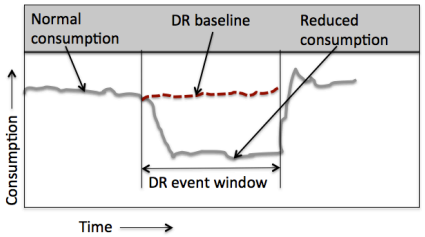
	Normal Consumption	DR Baseline	Reduced Consumption
Goal	Planning, DR	Curtailment calculation	Planning, DR, dynamic DR
Timing	Outside the DR event	Outside the DR event	During the DR event
Historical data	Readily available	Readily available	Sparse or non-existent
Compute requirements	Offline or real-time	Offline	Real-time for dynamic DR
Profile changes	Gradual	Gradual	Abrupt at DR event boundaries
Prior Work	Several	Several	**None**

We are the first to address this problem using data from DR experiments done on USC campus. **(Aman et. al., 2016), (Chelmis et. al., 2015)**

Key Challenges

- Unavailability of reduced consumption data
- Cancellation of DR event when found violating **thermal comfort limits** of occupants.
- Reduced consumption is affected by several factors
  - time of day/ day of week
  - reduction strategy
  - human behavior
  - external/environmental factors, e.g., temperature

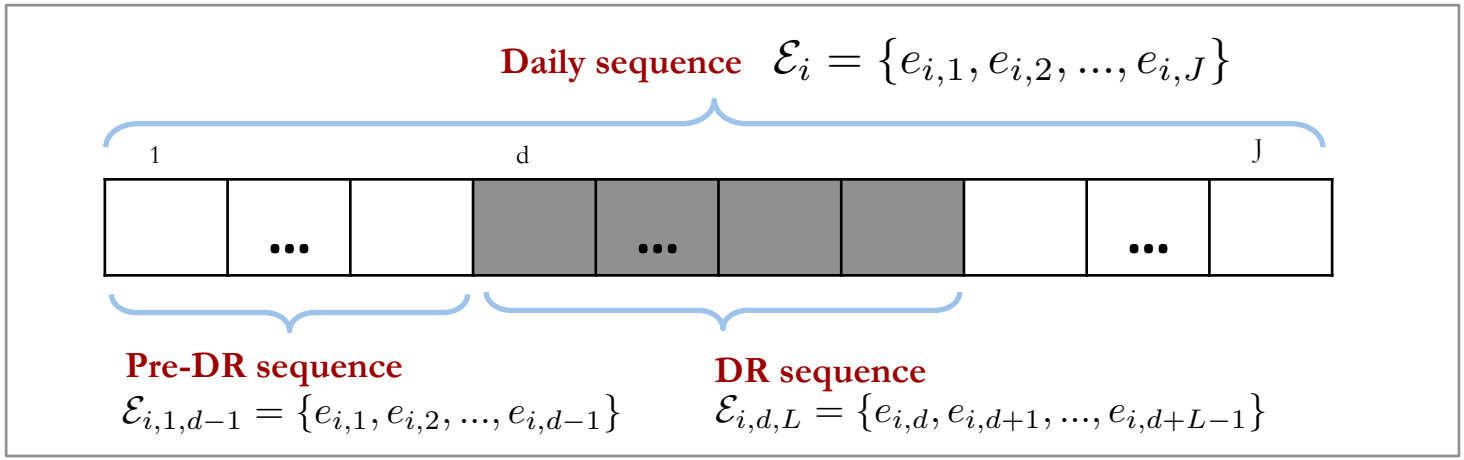
- Time series models that work well for normal consumption prediction are ineffective for reduced consumption prediction, due to
  - **abrupt changes** in consumption profile at the beginning and end of the DR event
  - **insufficient recent observations** within the DR window for a time series model to be trained reliably



Hypothesis

Historical data from the past DR events can be used as predictors for reduced consumption.

Consumption Sequences



$e_{i,j}$ – Electricity consumed on day *i* in interval *j*

$\mathcal{E}_{i,s,l}$ – Subsequence of daily sequence $\mathcal{E}_i$ starting at *s* of length *l*

L – Length of the DR interval

d – The interval when DR begins

J – Number of intervals in a day

$$d > 1$$

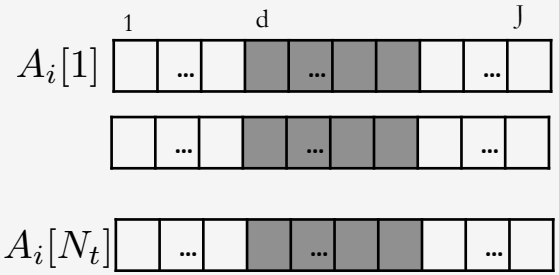
$$d + L - 1 \leq J$$

Contextual Attributes

- **Time Series attributes:** vary over intervals
  - temperature, dynamic pricing, occupancy, etc.
- **Static attributes:** same for all intervals
  - day of week, holiday, etc.

Time series attributes

N_t - # time series attributes



Static attributes

N_s - # of static attributes

$$B_i[1] \ldots B_i[N_s]$$

Daily Context

$$C_i = \langle A_i[1], \ldots, A_i[N_t], B_i[1], \ldots, B_i[N_s] \rangle$$

$$A_i[k] = \{a_{i,1}, a_{i,2}, \ldots, a_{i,J}\}$$

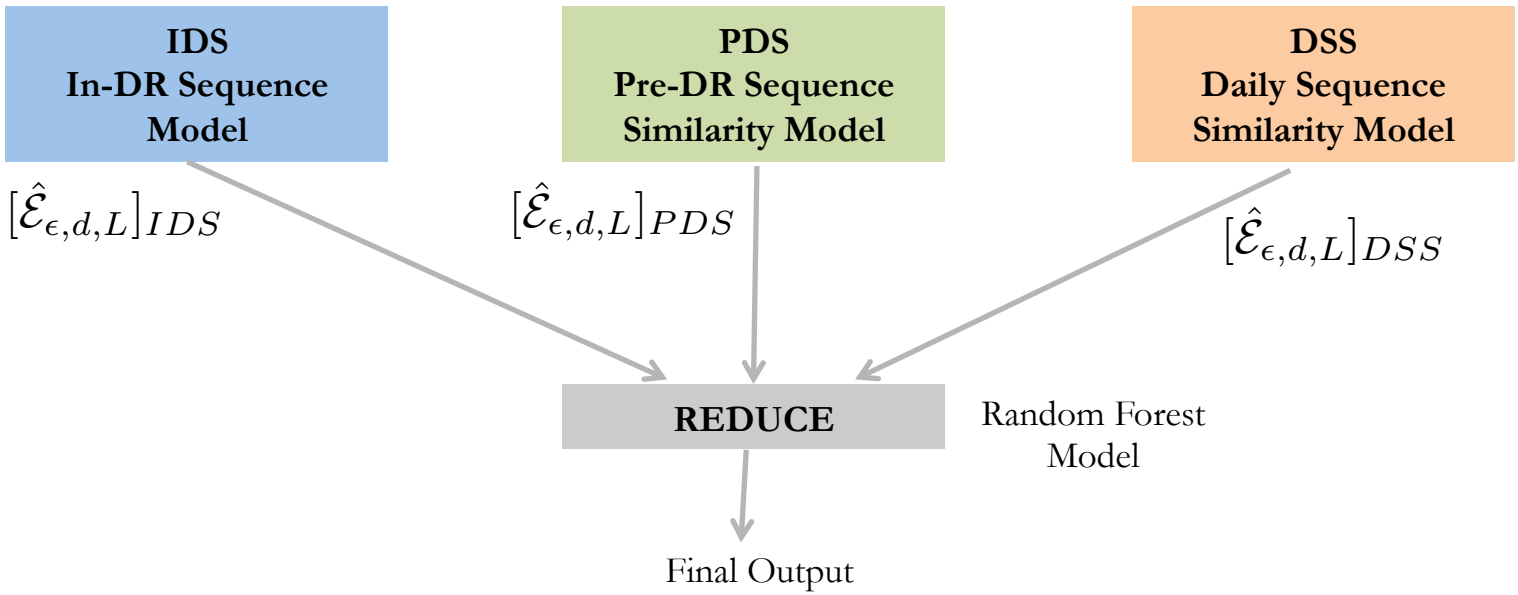
Pre-DR Context

$$C_{i,1,d-1} = \langle A_i[1], \ldots, A_i[N_t], B_i[1], \ldots, B_i[N_s] \rangle$$

$$A_i[k] = \{a_{i,1}, a_{i,2}, \ldots, a_{i,d-1}\}$$

Correspond to the Daily Sequence and Pre-DR Sequence defined previously.

REDUCE – Reduced Consumption Ensemble



- $[\hat{\mathcal{E}}_{\epsilon,d,L}]m$ – In-DR sequence predicted by model m on day ϵ
- Ensemble Models combine base models that model different behaviors, for e.g., **mean behavior, context dependent behavior**, etc.
- Random Forest Models are found to perform better than a single regression tree **(Breiman, 2001)**

IDS – In-DR Sequence Model

- Models "mean behavior"
- Similar to the averaging approach used by the utilities/ISOs to calculate the DR baseline.
- While utilities average over **similar** non-DR days, IDS averages over **all** DR days.

- Advantages:
  - Low computation cost – suitable for real-time predictions
  - Uni-variate model – low data collection cost

- Predicted sequence is given by:

$$[\hat{\mathcal{E}}_{i,d,L}]_{IDS} = \frac{1}{|\mathcal{E}|}\sum_{\epsilon=1}^{|\mathcal{E}|}\mathcal{E}_{\epsilon,d,L}$$

$\mathcal{E}$ is the set of historical DR days

PDS – Pre-DR Sequence Similarity Models

If two DR days have similar pre-DR sequences, their in-DR sequences would be similar.

- Pre-DR sequence
- Pre-DR context

} Used to select **similar** DR days

- Similarity is calculated by:

$$SimScore(\epsilon, i) = sim(\langle \mathcal{E}_{\epsilon,1,d-1}, \mathcal{C}_{\epsilon,1,d-1} \rangle, \langle \mathcal{E}_{i,1,d-1}, \mathcal{C}_{i,1,d-1} \rangle)$$

- Selected days are sorted based on decreasing similarity and weighed accordingly.
- Predicted sequence is given by:

$$[\hat{\mathcal{E}}_{i,d,L}]_{PDS} = \frac{1}{|\mathcal{E}|} \sum_{\epsilon=1}^{|\mathcal{E}|} \omega_\epsilon \times \mathcal{E}_{\epsilon,d,L}$$

$\mathcal{E}$ is the set of historical DR days
ω_ϵ is the weight on day ϵ

PDS models context dependent behavior

DSS – Daily Sequence Similarity Models

- Daily sequence
- Daily context

} Used to discover **clusters** of daily profiles

- Form daily profiles for each day

$$\mathcal{P}_\epsilon = \langle \mathcal{E}_\epsilon, \mathcal{C}_\epsilon \rangle$$

- Cluster daily profiles and let c_m be the centroid of each cluster

- Probability of a given DR day belonging to a cluster is given by:

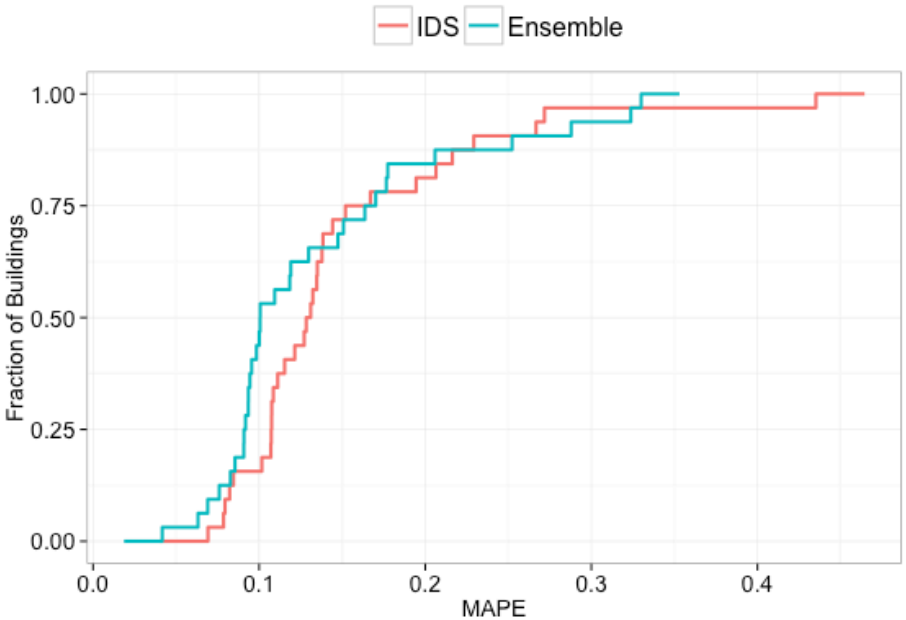
$$P(i \in C_m) = \frac{1}{\alpha \| \mathcal{P}_{i,1,d-1} - \mathcal{P}_{c_m,1,d-1} \|_2}$$

α is constant used to normalize the probability values between 0 and 1

- Predicted sequence is given by:

$$[\hat{\mathcal{E}}_{i,d,L}]_{DSS} = \frac{1}{N_k} \sum_{m=1}^{N_k} P(i \in C_m) \times \mathcal{E}_{c_m,d,L}$$

Results



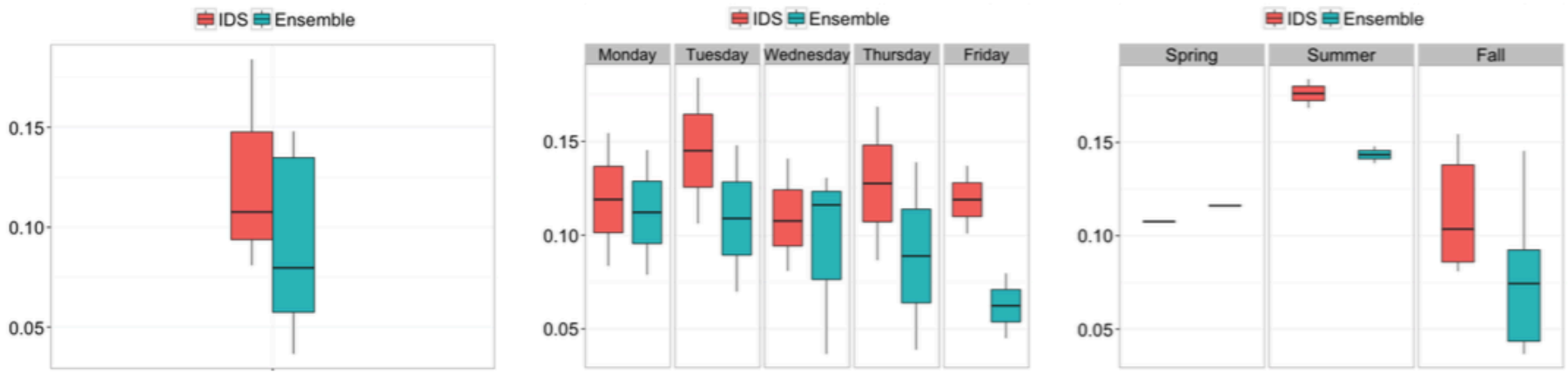
- 952 DR events (2012 – 2014)
- 32 USC buildings
- Contextual attributes:
  - temperature (NOAA)
  - day of week

- REDUCE outperforms the baseline IDS for about 70% of the buildings
- It also limits prediction error to <10% for over half the buildings
  – considered highly reliable by domain experts **(Aman et. al., 2015)**
- Overall average error is 13.5%, an improvement of 8.8% over the baseline

Scheduled Vs. Non-scheduled

- Scheduled – activities governed by schedules, for e.g., classrooms
- For non-scheduled:
  - REDUCE gives superior performance
  - IDS does not perform well due to the absence of repetitive human activity coupled to class schedules

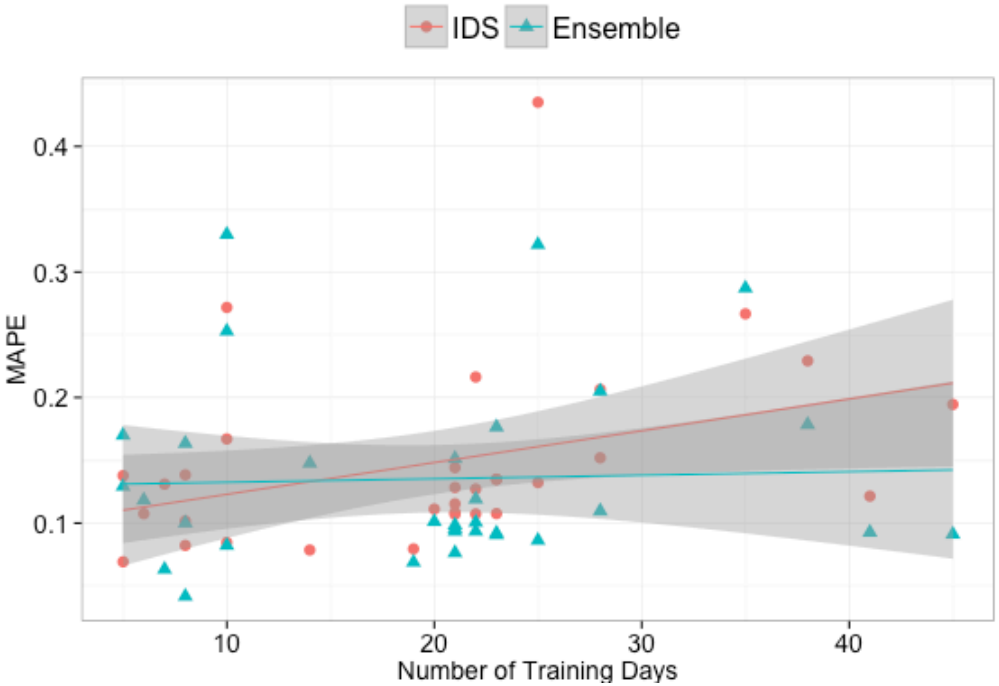
MAPE errors for Non-scheduled building (large computer labs, faculty and student offices)



(a) For all test days (b) By day of week (c) By seasons

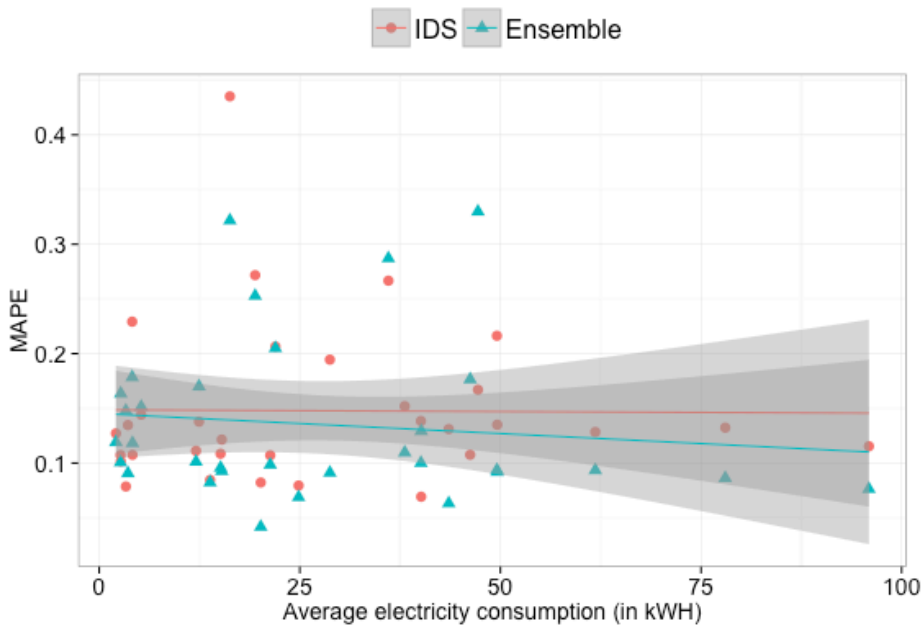
- **Corollary**: REDUCE would perform better for residential buildings (with non-scheduled activities).

Effect of Training Data



- REDUCE is not sensitive to the training data size -> reduces computational and storage requirements
- **Corollary:** REDUCE would allow accurate predictions to be made for new buildings with fewer historical data

Building Size



- For REDUCE, error decreases with increasing average consumption
  - more stable and predictable behavior for larger buildings
- Insight: The performance of REDUCE slightly improves for larger buildings

Contribution 3

Prediction Evaluation Measures

Common Evaluation Measures

o_i – observed value at interval *i*
p_i – predicted value at interval *i*
n – number of intervals for which prediction is made

Mean Absolute Percentage Error (MAPE)
Mean absolute Error (MAE) normalized by the observed value.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|p_i - o_i|}{o_i}$$

Coefficient of Variation of Root Mean Square Error (CV-RMSE)
Root Mean Square Error (RMSE) normalized by the mean of observed values

$$CVRMSE = \frac{1}{\bar{o}}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2}$$

Limitations of Current Evaluation Approaches

Dimension	Uni-dimensional focus on error measures
Prediction Bias	Insensitive to prediction bias Under-prediction is deleterious for estimating peak demand
Scale	Scale-dependent measures are unsuitable for comparing customers of different sizes
Reliability	Don't consider the frequency with which a model does good predictions: – # times a model outperforms the baseline – # times a model's error is within a tolerance level
Cost	Don't consider the costs – data collection – building models and running them
Application-relevance	Based on "abstract metrics"; not relevant to the end application

We propose holistic evaluation measures to address these limitations.

‘Application-specific’ Bias based Measure

o_i – observed value at interval *i*
p_i – predicted value at interval *i*
n – number of intervals for which prediction is made
α, β – penalty parameters associated with over- and under- predictions

Domain Bias Percentage Error (DBPE)

$$DBPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathcal{L}(p_i, o_i)}{o_i}$$

$$\mathcal{L}(p_i, o_i) = \begin{cases} \alpha \cdot |p_i - o_i|, & \text{if } p_i > o_i \\ 0, & \text{if } p_i = o_i \\ \beta \cdot |p_i - o_i|, & \text{if } p_i < o_i \end{cases}$$

over-prediction

under-prediction

Penalty parameters are configured for specific applications in consultation with the domain experts.

Reliability Measure

o_i – observed value at interval *i*
p_i – candidate model predicted value at interval *i*
b_i – baseline model predicted value at interval *i*
n – number of intervals for which prediction is made

Relative Improvement (RIM)
Fraction of predictions made by a candidate model better than the baseline model.

$$RIM = \frac{1}{n}\sum_{i=1}^{n} C(p_i, o_i, b_i)$$

$$C(p_i, o_i, b_i) = \begin{cases} 1, & \text{if } |p_i - o_i| < |b_i - o_i| \\ 0, & \text{if } |p_i - o_i| = |b_i - o_i| \\ -1, & \text{if } |p_i - o_i| > |b_i - o_i| \end{cases}$$

Candidate model performs better than the baseline

‘Application-specific’ Reliability Measure

o_i – observed value at interval *i*
p_i – candidate model predicted value at interval *i*
e_t – error threshold
n – number of intervals for which prediction is made

Reliability Threshold Estimate (REL)
Measures how frequently the errors fall within a set threshold

$$REL = \frac{1}{n}\sum_{i=1}^{n} C(p_i, o_i)$$

$$C(p_i, o_i) = \begin{cases} 1, & \text{if } \frac{|p_i - o_i|}{o_i} < e_t \\ 0, & \text{if } \frac{|p_i - o_i|}{o_i} = e_t \\ -1, & \text{if } \frac{|p_i - o_i|}{o_i} > e_t \end{cases}$$

The error threshold is set for specific applications in consultation with the domain experts.

Cost Measures

Computation Cost (CC)

application-independent

$$CC = CC_t + CC_p$$

CC_t – time taken to train a model

CC_p – time taken to predict using the trained model

Total Computation Cost (TCC)

application-specific

$$TCC = CC_t \cdot \tau + CC_p \cdot \pi$$

τ – number of times a model is trained in a duration of interest

π – number of times a model makes prediction in that duration

Cost-Benefit Measure

For prediction models using "big data", it is critical to consider the cost of building and using a model relative to the gain it provides.

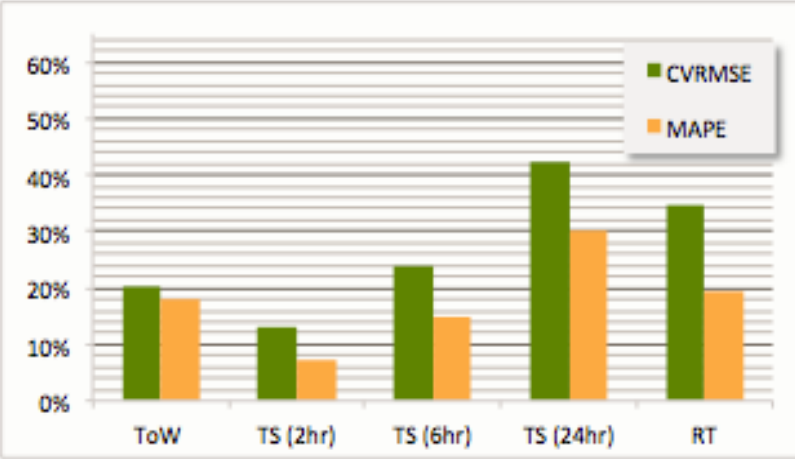
Cost-Benefit Measure (CBM)

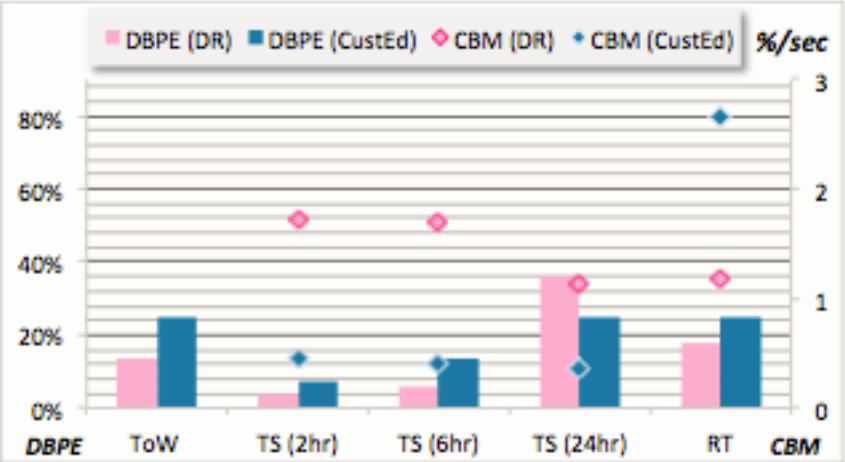
$$CBM = \frac{(1 - DBPE)}{TCC}$$

A model with high accuracy but with prohibitive cost may be unsuitable.

Results – Bias based DBPE measure

To avoid missing peaks, we favor over-predictions to under-predictions.
We set $\alpha = 0.5$ and $\beta = 1.5$ for DBPE





DBPE is uniformly smaller than MAPE.
CBM is lower for Regression Tree model due to high data collection cost for different features.

Conclusion

Conclusion

- We proposed **Dynamic Demand Response (D^2R)**
  - as a novel extension of the state-of-the-art DR practice in smart grids
  - as a prime example of **dynamic decision making** in smart grids

- We made following contributions:
  - Proposed novel model for prediction with partial data
  - Proposed novel ensemble model for prediction of reduced consumption
  - Proposed holistic measures to evaluate prediction performance

- Our proposed models are being used (or in process of deployment) at the USC Facilities Management and will eventually be used at the LA Department of Water and Power (LADWP).

Publications

- Aman et. al., Holistic Measures for Evaluating Prediction Models in Smart Grids, Transactions in Knowledge and Data Engineering, 27(2), 2015.
- Aman et. al., Influence-driven Model for Time Series Prediction from Partial Observations, AAAI 2015.
- Aman et. al., Learning to REDUCE: A Reduced Electricity Consumption Prediction Ensemble, AAAI Workshop, 2016.
- Aman et. al., "Prediction Models for Dynamic Demand Response: Requirements, Challenges, and Insights", IEEE SmartGridComm, 2015.
- Aman et. al., Addressing Data Veracity in Big Data Applications, Poster Paper, IEEE Conference on Big Data, 2014.
- Aman et. al., Energy Management Systems: State of the Art and Emerging Trends, IEEE Communications Magazine, 2013.
- Aman, Analytics for Demand Response Optimization in a MicroGrid, SDM Doctoral Forum, 2012
- Aman et. al., Improving Energy Use Forecast for Campus Micro-grids using Indirect Indicators, ICDM International Workshop on Domain Driven Data Mining, 2011.

Publications

- Frincu et. al., Enabling Automated Dynamic Demand Response: From Theory to Practice, ACM International Conference on Future Energy Systems (e-Energy), 2015.
- Pal et. al., On Online Time Series Clustering For Demand Response: OPTIC - A Theory to Break the 'Curse of Dimensionality', ACM Conference on Future Energy Systems, 2015.
- Chelmis et. al., Estimating Reduced Consumption for Dynamic Demand Response, AAAI Workshop on Computational Sustainability, 2015.
- Simmhan et. al., Cloud-based Software Platform for Data-Driven Smart Grid Management, IEEE/AIP Computing in Science and Engineering, Jul/Aug 2013.
- Simmhan et. al., Towards Data-driven Demand-Response Optimization in a Campus Microgrid, ACM Workshop On Embedded Sensing Systems For Energy-Efficiency In Buildings, 2011.
- Simmhan et. al., Adaptive Energy Forecasting and Information Diffusion for Smart Power Grids, IEEE International Scalable Computing Challenge, 2012.

References

- Ziekow et. al., "The potential of smart home sensors in forecasting household electricity demand", IEEE SmartGridComm, 2013.
- Meek et. al., Autoregressive tree models for time-series analysis, SIAM Conference on Data Mining (SDM) 2002.
- L. Breiman. Random forests. Machine learning, 45(1):5–32, 2001.
- Arnold et. al., Temporal causal modeling with graphical granger methods. In International conference on Knowledge discovery and data mining (KDD '07)
- Tibshirani, Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58(1), 1996.
- Cuevas-Tello, J. C. et. al., Uncovering delayed patterns in noisy and irregularly sampled time series: an astronomy application. Pattern Recognition 43(3), 2010.
- Kreindler, D. M. et al., The effects of the irregular sample and missing data in time series analysis. Nonlinear dynamics, psychology, and life sciences 10(2), 2006.

Thank you!