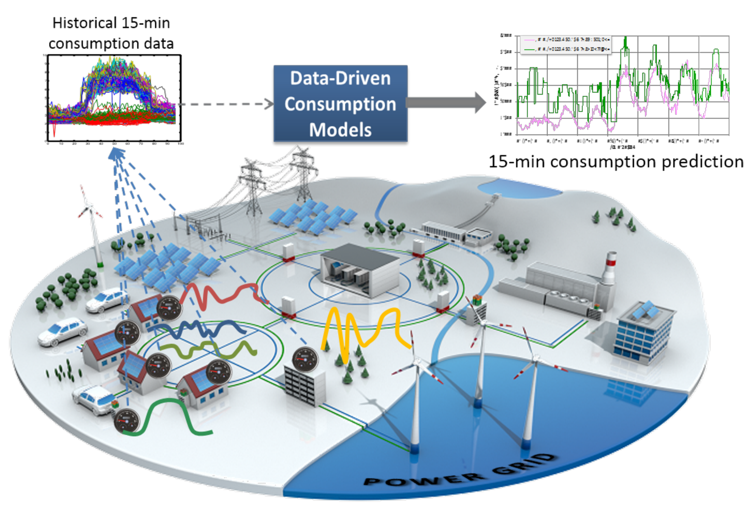
# Influence-driven Model for Time Series Prediction from Partial Observations

**Saima Aman, Charalampos Chelmis, Viktor Prasanna**

## Motivation: Smart Power Grids

- Real-time monitoring
- Advanced Big Data Analytics

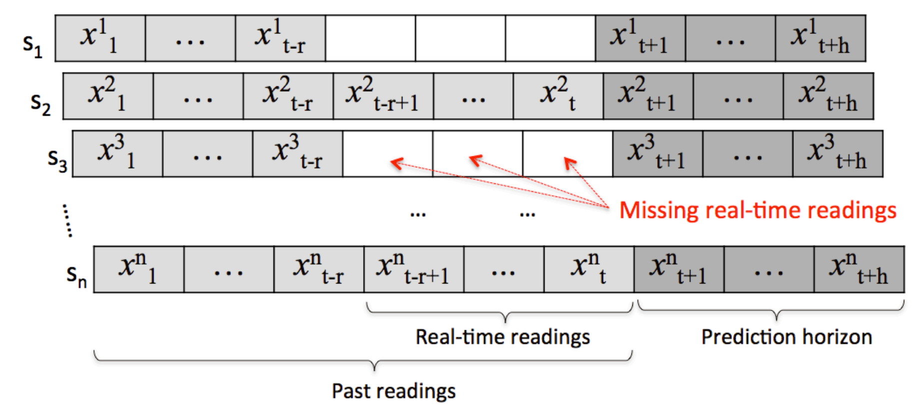



### Problem

- Bandwidth limitations of existing communication networks
- Transmission limited by consumers due to security and privacy concerns
- Only partial real-time data available at the central node

### Challenges

- Millions of sensors used
- High-resolution predictions required for all sensors
- Unavailability of real-time data for most sensors
- Effective prediction horizon becomes large




### Current Approaches

- Data compression and aggregation
- Estimate missing data using interpolation
- These approaches ineffective for partial data

| Missing Data | Partial Data |
|---|---|
| arbitrary unavailability of data | systematic unavailability of data |
| for variable durations | for known durations |
| from unknown sensors | from known sensors |
| diverse factors, such as transmission failure | due to non-transmission of data |
| data never available | data available later on |

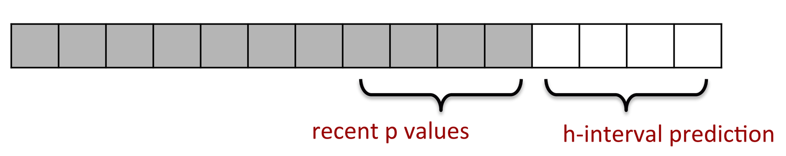
### Solution

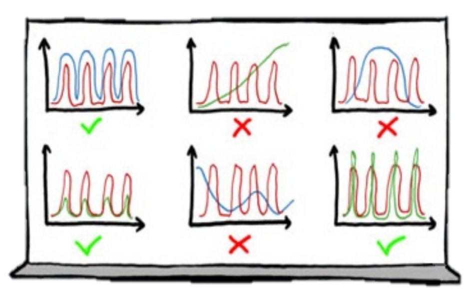
Reduce the transmission load: use real-time data from only a small subset of sensors

## Prediction Models using Partial Data

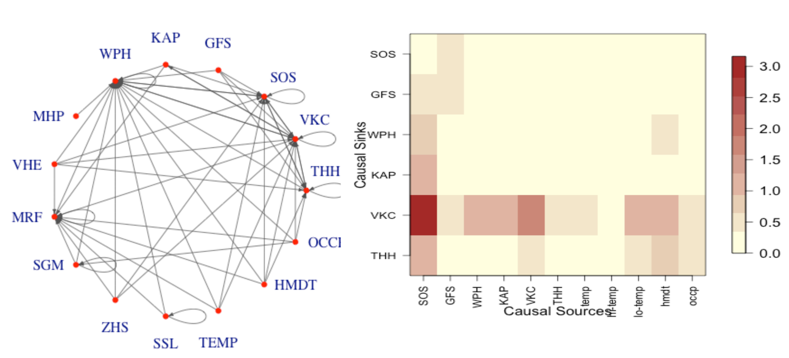
### Baseline model (ART)

Auto-regressive tree ART($p$, $h$) – uses recent $p$ values of a variable as features in a regression tree for $h$ interval ahead prediction

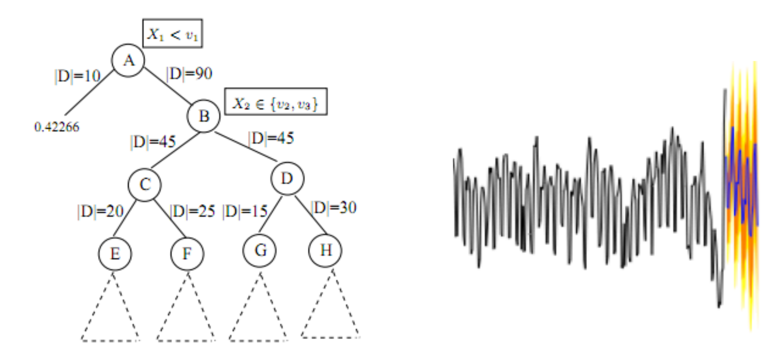


1. Identify dependencies between time series from recent historical data



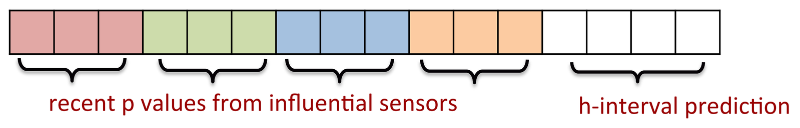
2. Discover influential sensors using lasso based method



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

### Influence Model (IM)

More recent real-time values of other sensors become more useful as predictors than a sensor's own relatively older values (used in ART).




| ART | IM | LIM | GIM |
|---|---|---|---|
| • Use a sensor's own past values for prediction | • Use influential sensors' past values for prediction | • Use influential sensors' past values for prediction<br>• Sensors selected **locally** for each sensor | • Use influential sensors' past values for prediction<br>• Sensors selected **globally** for all sensors |

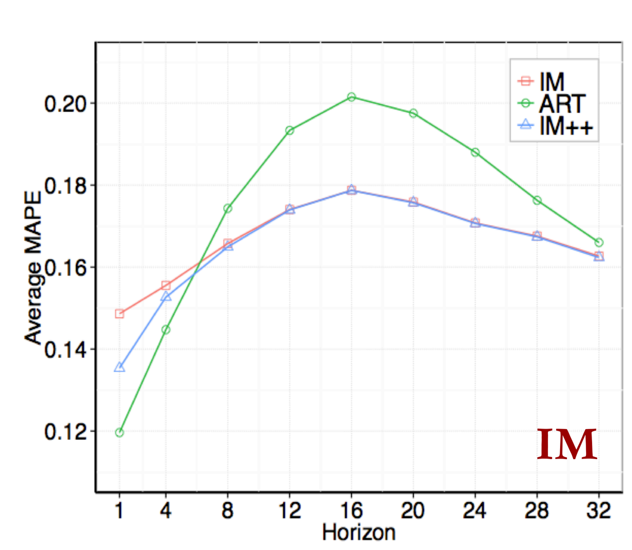
**Decreasing number of sensors required to transmit data in real time**

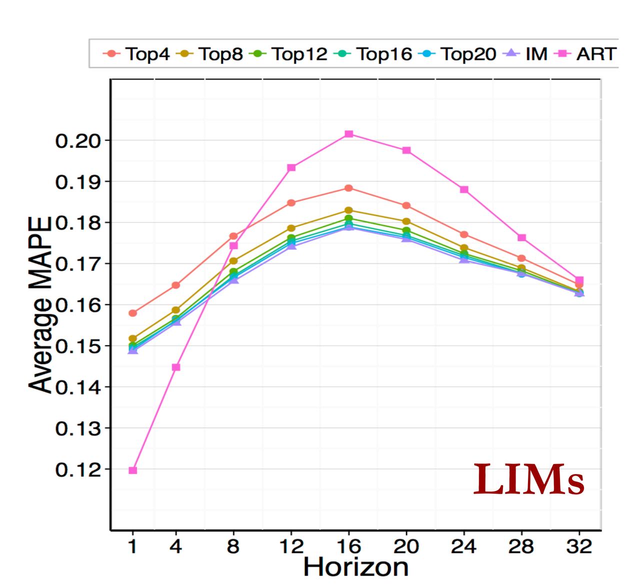
## Experiments and Results

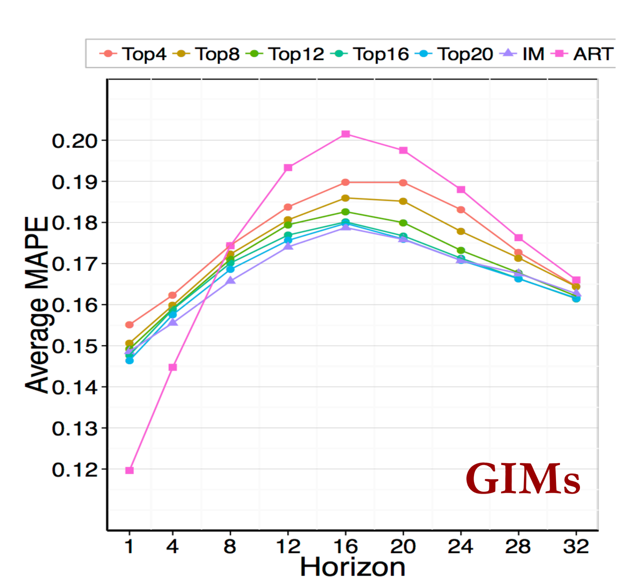
### Data

**Smart Meters**
170 meters in USC
15-min observations

**Weather**
Downtown LA/ USC station
Hourly observations




### Performance Highlights

- IM reduces MAPE by up to 10%
- Beyond 2 hours, IM, LIMs and GIMs outperform ART, which suffers from longer effective horizons
- GIM uses the same set of influential sensors to make predictions for all sensors and still achieves comparable performance.
- Only ≈0.5% increase in prediction error over ART while using just ≈7% of smart meters.

### Conclusions

- Partial data problem encountered in many sustainability domains that use sensor data
- Standard models for short term predictions perform poorly when trying to predict using partial data
- Our proposed influence models make predictions using real time data from only a few influential sensors, while still achieving comparable or better performance than the baseline

ACKNOWLEDGEMENT

Los Angeles Department of Water & Power

CONTACT
saman@usc.edu
chelmis@usc.edu
prasanna@usc.edu

USC School of Engineering

University of Southern California