Estimating the behind-the-meter solar generation with existing infrastructure

Emre Can Kara,
Michaelangelo Tabone, Sila Kiliccote

AI-at-SLAC

July 11, 2017
Our team

Science Directorate > Applied Energy Program:
- Materials for energy
- Subsurface Technology
- 21st Century Electric Grid

Team of 15 researchers, scientists and engineers.

Major funding:
- DOE EERE
- CEC EPIC
- ARPA-e

gismo.slac.stanford.edu
What do we do @GISMo?

**POWER GRID**
- Distributed control
- Data-driven planning and operations with DERs
- Pricing mechanisms and market structures
- Business models
- Regulatory frameworks

**AMBIENT / BUILDING INTELLIGENCE**
- Interaction design
- Seamless integration
- Embedded
- Context aware
- Personalized
- Adaptive
- Anticipatory

**MOBILITY**
- Mobility as a service
- Electrification of transportation
- Data-driven modeling and analysis
- Vehicle-to-building
- Vehicle-to-grid
Integrating Renewable Energy

Transmission Grid

Distribution Grid

Supply = Demand
Integrating Renewable Energy

Transmission Grid

Distribution Grid

Supply = Demand
VADER: Visualization and Analytics for Distributed Energy Resources

Enabling high penetration of renewables requires:

- More comprehensive and dynamic situational awareness; and
- Active control of the distribution network.

Traditional approach/technology is inadequate:

- Partial observability
- New devices and models
- Heterogenous data sources and data streams
Plenty of data that is not being used!

B2G – Building to Grid
μgrids – Microgrids
V2B – Vehicle to Building
VGI – Vehicle to Grid

Synchrophasors
Micro-synchrophasors
Line sensors

Power System
Buildings & DERs
Mobility

Interval meters
Smart meters
Trend logs from building management Systems (BMS)
Charging sessions
Driving patterns
VADER Introduction

Visualization and Analytics of Distributed Energy Resources

Raw Data  Access  Ingestion  Analytics  Applications

Utility  Data Plug  Virtual SCADA  What Now  Services Visualization Reporting
Public
3rd Party

Analytics

What If

Traditional: SE/PF
Direct: ML/Stat

Funded by:

SunShot
U.S. Department of Energy
VADER Analytics

VADER key analytics pieces:
- Solar disaggregation
- Switch configuration
- ML-based power flow analysis
- Network topology detection
- Load forecasting
VADER Infrastructure
VADER Infrastructure

- Polyglot persistence:
  - Why?
Introduction

▪ Increasing solar penetration
  – Behind-the-meter
  – Distribution-level
▪ Switches maintain a radial structure
▪ Load is masked
▪ Visibility into behind-the-meter solar generation is limited

How do we gain more visibility into the load and solar generation?
Introduction

- Traditional approaches include:
  - Physics-based models:
    - Geometry of the array
    - Nameplate capacity
    - Site-specific irradiance measurements (diffuse horizontal, direct normal, and global horizontal irradiance)
      - Costly to obtain DHI and DNI often only GHI is available

Can we use existing measurements that are collected by the utilities to estimate behind-the-meter solar in real-time? If so, how accurate these estimates would be?

- Could be costly to do in real-time.
Problem formulation

- Contextually supervised source separation*: 
  - Observing an aggregate measurement of signals of interest
  - Each signal can be represented as a linear model
  - Contextual knowledge about signal characteristics
  - Used in Non-intrusive Load Monitoring

Error between the estimates of the signal and what the linear model would do

\[
\text{minimize } \sum_{Y_i, \Theta_i} \left\{ \alpha_i \ell_i ((Y_i - X_i \Theta_i)) + \eta_i g_i(Y_i) + \gamma_i h_i(\Theta_i) \right\}
\]

subject to \( Y_{agg} = \sum_{i=0}^{L} Y_i \)

Problem formulation

- **Real-time SCADA measurements** (typically 4 seconds)
- **AMI**: overnight updates
  (1-min to 15-min sampling rate)
Problem formulation

Day-ahead training problem (learning the model):

**Inputs**
- AMI data
- $NL_i$
- Sparse # of solar sensors
- Outdoor Temp.

**Outputs**
- Load and solar at each meter
- $S_i$, $L_i$

\[ \text{AMI data} \quad NL_i \quad \text{Sparse # of solar sensors} \quad \text{Outdoor Temp.} \quad \text{Load and solar at each meter} \]
Problem formulation

\[
\begin{align*}
\text{minimize} & \quad \alpha_{AL} \|AL - \theta_{AL}X_{AL}\|_2 + \sum_{i=1}^{N} \alpha_i \|S_i - \theta_iX_i\|_2 \\
\text{subject to} & \quad L_i + S_i = NL_i, \quad \forall i \\
& \quad S_i \leq 0 \quad \forall i, \\
& \quad L_i \geq 0 \quad \forall i, \\
& \quad \sum_{i=1}^{N} L_i = AL \\
& \quad \theta_i \geq 0.
\end{align*}
\]

Aggregate load is much more predictable

Need to be tuned
Problem formulation

Creation of aggregate load regressors:

\[
T_t^{(LB)} = \begin{cases} 
0 & T_t < LB \\
10 & T_t > LB + 10 \\
T_t - LB & LB \leq T_t \leq LB + 10 
\end{cases}
\]

<table>
<thead>
<tr>
<th>Temp_t</th>
<th>Int</th>
<th>( T_t^{(60)} )</th>
<th>( T_t^{(70)} )</th>
<th>( T_t^{(80)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>66</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>68</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>70</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>72</td>
<td>1</td>
<td>10</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>74</td>
<td>1</td>
<td>10</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>80</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>82</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>84</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>4</td>
</tr>
</tbody>
</table>
Tuning alphas

- Intuitive choice: the inverse of the expected variance of model's errors
- During nighttime there is no solar generation
  - Net load = Aggregate Load
  - Estimate the error for the AL model
- During day time we can estimate the total error.
- Assuming the all errors are independent:

\[ \text{Var}(\epsilon_{AL}) = \text{Var}(\epsilon_{Total}) - \sum_{i=1}^{N} \text{Var}(\epsilon_i) \]
Problem formulation

Real-time estimation problem:

**Input**
- Data from load aggregation point (substation)
- Solar from sparse # of sensors
- Outdoor Temp.

**Outputs**
- Solar at each meter
Problem formulation

minimize  \( \alpha_{AL} \| AL - \theta_{AL} X_{AL} \|_2 + \sum_{i=1}^{N} \alpha_i \| S_i - \theta_i X_i \|_2 \) 

subject to  \( AL + \sum_{i=1}^{N} S_i = \sum_{i=1}^{N} NL_i \),

Observed through SCADA

\( S_i \leq 0 \quad \forall i. \)
Dataset

Subset of Pecan Street dataset:

- Load and solar sub-metered at 1-minute resolution
- 110 households in TX
- Total duration 7-days
- Divided into:
  - 55 homes with solar
  - 55 homes without solar
  - Varying number of distribution-level solar (used as proxy)
Results

- We use coefficient of variation for homes with solar.
- We report only RMSE for homes without solar.

\[
CV = \frac{\sqrt{\sum_{i=1}^{T} (truth_t - estimate_t)^2}}{\frac{1}{T} \sum_{i=1}^{T} truth_t}
\]

\[
= \frac{RMSE}{Mean \ truth}
\]
Training Results (Untuned vs. Tuned)

Home # 2638
- CV: 0.18

Home # 3723
- RMSE: 0.48

Home # 7982
- CV: 0.17

Aggregate
- CV: 0.46

Home # 1268
- CV: 0.17
- RMSE: 0.06

Home # 2982
- CV: 0.09

Aggregate
- CV: 0.3
Real-time Results (Untuned vs. Tuned)

Home # 2638
CV: 0.42
RMSE: 0.29

Home # 3723
CV: 0.42

Home # 7982
CV: 0.42

Aggregate
CV: 0.34

Power, kW
0 1 2 3 4
Hour of Day
Disagg. Solar, True Solar
Conclusions & Future Work

▪ Useful tool to estimate behind-the-meter solar in real-time.
  – Can also be used to select the optimal number of proxies and their locations for a guaranteed performance of estimation.
▪ Incorporate losses in the network
▪ Expanding the disaggregation strategy to incorporate storage
▪ Running the model on a larger network in SCE’s territory through our partnership in VADER.
Thank you!

Michaelangelo Tabone
Post-doctoral Fellow
Stanford University,
SLAC National Accelerator Laboratory
mtabone@stanford.edu

Emre Can Kara
Associate Staff Scientist
SLAC National Accelerator Laboratory
ekara@stanford.edu

Sila Kiliccote
Staff Scientist
SLAC National Accelerator Laboratory
silak@stanford.edu
Additional Slides

Predictions of BTM Solar (homes with solar)

Predictions of BTM Solar (homes without solar)

Predictions of Aggregate Load

Number of Proxies

flat tuned