Fine-grained Photovoltaic Output Prediction using a Bayesian Ensemble

Prithwish Chakraborty[1,2] , Manish Marwah[3], Martin Arlitt[3], and Naren Ramakrishnan[1,2]

[1]Department of Computer Science, Virginia Tech, Blacksburg, VA 24061
[2]Discovery Analytics Center, Virginia Tech, Blacksburg, VA 24061
[3]Sustainable Ecosystems Research Group, HP Labs, Palo Alto, CA 94304

Introduction : Solar PV importance

- Increasing emphasis on local and distributed power generation from renewable energy resources
- Reduction in carbon footprint.
- Lower transmission and distribution losses through local generation.
- Other socio-economic impetus.

Introduction : Solar PV challenges

- Photovoltaic (PV) array generation highly variable and intermittent
- PV output has high temporal dependency
- Also depends on environmental factors such as cloud cover, temperature
- Effective usage requires a near-optimal prediction of PV array output – can be used to match workload profile.

Introduction : Problem Definition

- Available data on *j-th* hour of the *i-th* day :
  - Historic PV generation
  - Historic weather conditions
  - Forecast for future weather conditions

- Goal :
  - Predict Solar output for the hours $j + 1, j + 2, \ldots, 24$ of the i -*th* day
  - Revise predictions as the hour increases i.e. new data is available

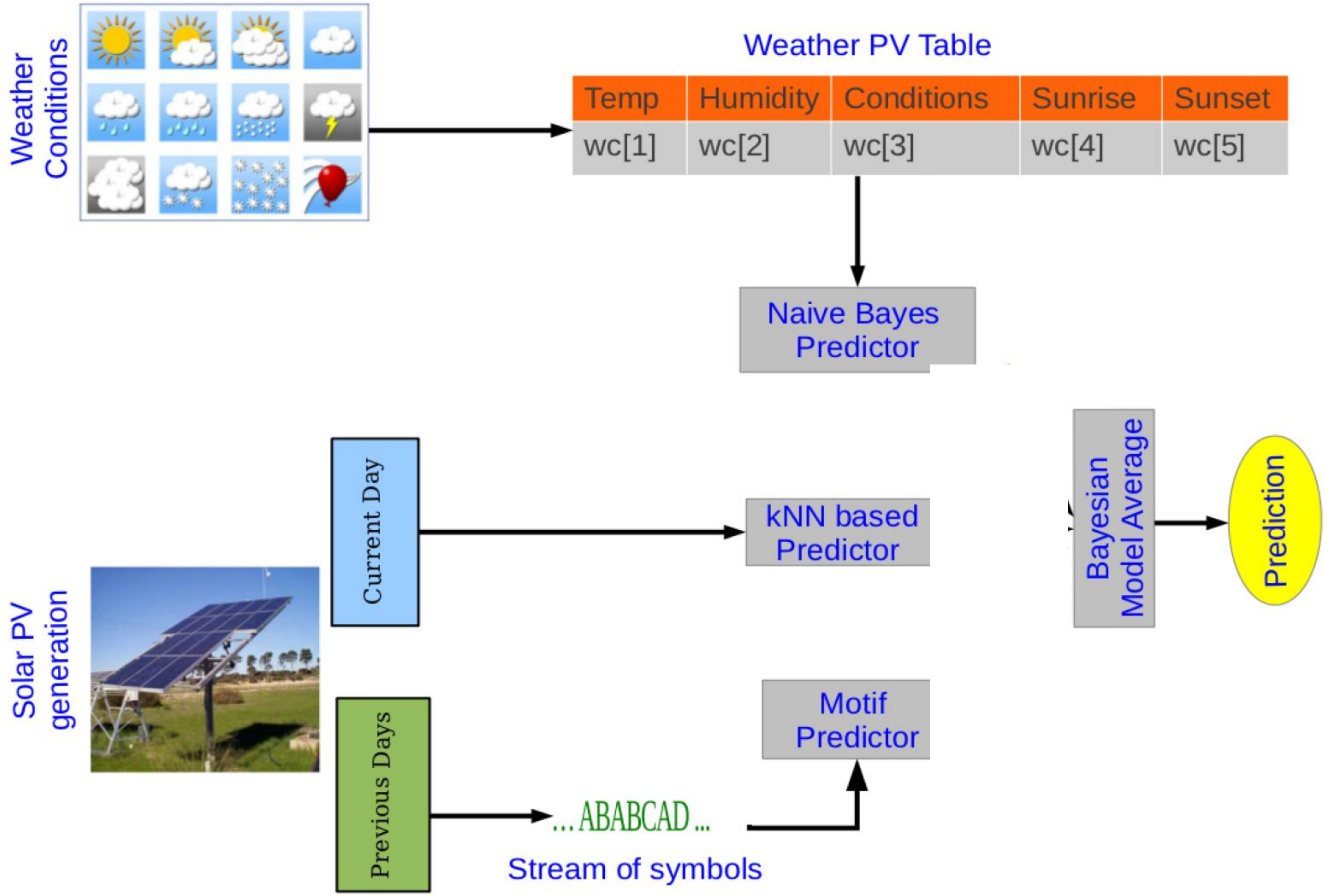
Brief Summary of this work

- Hourly prediction of Solar PV array output with a prediction bound of a day.
- Uses inherent temporal dependency of PV output to get a global trend of PV output : time series motif used.
- Local variations modeled by the corresponding environmental conditions and output from observed values of PV output for the past hours of the day.
- Use of a general purpose Bayesian Ensemble to combine the global trend and local variations to get the final prediction.

Existing Research

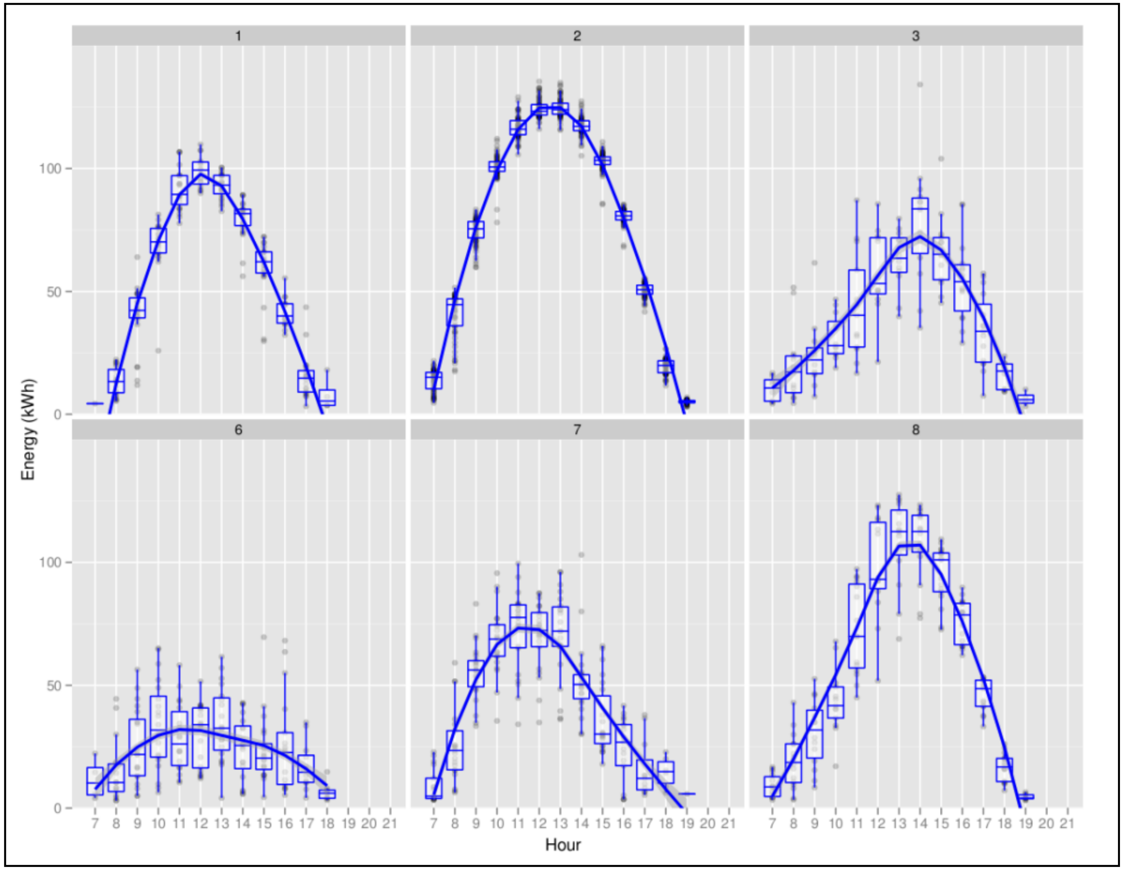
- Classical time series predictions such as ARMA and ARIMA.
- Weighted average methods for PV output prediction : prediction horizon is generally one hour.
  - Cox, 1961
  - Piorno et al. , 2009
- Prediction of irradiation pattern : close correspondence between irradiance and Solar PV values.
- Bofinger et al. (2006) : Global forecasts of an European weather center modified by local statistical models.

Proposed Framework



Bayesian Ensemble: Preprocessing

- Extraction of profiles from training data
  - Solar PV values for each day represented by a vector where the dimensions correspond to the hour of the day
  - K-means clustering on the training vectors to get a set of profiles of pre-determined cardinality.



Bayesian Ensemble : Motif Discovery

- Frequent episode counting to get motifs
  - Stream of days converted to stream of profile labels.
  - Start with a pre-determined window size *W*.
  - Slide the window over the data stream and count eligible episodes

Definition : Eligible Episodes

For a window W_i ($|W_i| = W$) containing labels $d_i =< d_{i-W,} \quad \ldots, d_{i-1} >$, eligible episodes are defined as all such sequences $ep =< d_{p_1}, d_{p_2}, d_{p_3} \ldots >$, such that $p_1 < p_2 < p_3$

Bayesian Ensemble : Motif Discovery

- Count maximally frequent eligible episodes using Apriori algorithm (Agarwal and Srikant, 1994)

Definition : Maximally frequent Eligible Episodes

An eligible episode is maximally frequent iff :

1. $sup_{ep_i} > \tau$
2. $\nexists\ ep_i$ such that $ep_i < ep_j$

Bayesian Ensemble : Motif based Classification

- While predicting for *i-th* day, find motifs that can contain the *i-th* label and labels of some of the previous days in a window size $W - 1$

- Set of such motifs denoted by $\langle ep(D_n) \rangle = \{ep(D_n)_1, ep(D_n)_2, \ldots, ep(D_n)_p\}$ and support counted as
$$\mathrm{sup}(\langle ep(\mathrm{D_n}) \rangle) = \sum_p sup_{ep(D_n)_p}$$

Classifier Output : Motif Based Classifier

$$Pr(\vec{x}_i \in D_n | \vec{x}_{i-1}, \vec{x}_{i-2}, .., \vec{x}_{i-W-1}, C_3) = \frac{sup(\langle ep(D_n) \rangle)}{\sum_{n=1}^{N} sup(\langle ep(D_n) \rangle)}$$

Bayesian Ensemble : KNN based Classification

- Find Euclidean distance between partially observed current day and corresponding partial profile centroids.
- Take normalized inverse distance as classifier output

Classifier Output : k-NN based classifier

$$Pr(\vec{x_i} \in D_n | \vec{x}_i(1:j), C_2) = \frac{1}{\phi \| \vec{x}_i(1:j) - \vec{\mu}_n(1:j) \|_2}$$

where $\phi = \sum_n \frac{1}{\|\bar{x}_i(1:j) - \bar{\mu}_n(1:j)\|}$

Bayesian Ensemble : Naïve Bayes based Classifier

- Build table of actual weather conditions (γ) and corresponding PV profile label of the day

- Weather forecast denoted by

$$\rho_{i,j} = \langle \rho_{i,j+1,1}, \rho_{i,j+2,2}, \dots \rho_{i,J,J-j} \rangle$$

- Posterior Probability calculated as:

$$\Pr(D_n | \rho_{i,j+t,t}, \gamma) \propto \left(\prod_k L(D_n | \rho_{i,j+t,t}[k]) \right) \Pr(D_n)$$

Classifier Output : Naïve Bayes based classifier

$$Pr(D_n | \rho_{i,j}, \gamma, C_1) = \frac{\sum_{t=1}^{J-j} Pr(D_n | \rho_{i,j+t,t}, \gamma)}{\sum_{n=1}^{N} \sum_{t=1}^{J-j} Pr(D_n | \rho_{i,j+t,t}, \gamma)}$$

Bayesian Ensemble : BMA

- Probabilities of profiles combined using Bayesian Model averaging

Ensemble Output : BMA probability

$$
\begin{aligned}
P(\vec{x_i} \in D_n | D_{i,j} = \langle \vec{x}_i(1{:}j), \rho_{i,j}, \vec{x}_{i-1}, \vec{x}_{i-2}, \ldots, \vec{x}_{i-W-1} \rangle) = {} & P(\vec{x_i} \in D_n | C_1, \rho_{i,j}) P(C_1 | D_{i,j}) + \\
& P((\vec{x_i} \in D_n | C_2, \vec{x}_i(1{:}j)) P(C_2 | D_{i,j}) + \\
& P(\vec{x_i} \in D_n | C_3, \vec{x}_{i-1}, \vec{x}_{i-2}, \ldots, \vec{x}_{i-W-1}) P(C_3 | D_{i,j})
\end{aligned}
$$

Bayesian Ensemble : BMA

- Assume uniform priors for classifiers

- Final Prediction output through estimation

Ensemble Output : BMA estiamtion

$$E(\vec{x}_i(j+1:J)|D_{i,j}) = \sum_n \vec{\mu}_n(j+1:J) P(\vec{x}_i \in D_n | D_{i,j})$$

Experimental Observations

- Datasets
  - 154 kW PV installation at a commercial building in Palo Alto, CA, USA.
    - 5 min interval data
    - March 2011 to November 2011 (267 days)
    - Hourly weather data from a nearby weather station
    - Weather data : temperature, humidity, visibility and weather conditions related to cloud cover/precipitation.
  - Solar and weather data from Amherst, MA

Experimental Observations : Parameter Estiamtion

- Under uniform prior assumption,

$$P(C_l|\mathcal{D}_{i,j}) \propto P(\mathcal{D}_{i,j}|C_l)$$

- Heuristics used to compute the priors:

$$\begin{aligned}
P(\mathcal{D}_{i,j}|C_3) &= \theta \\
P(\mathcal{D}_{i,j}|C_2) &= min(1-\theta, \alpha \times j + \beta) \\
P(\mathcal{D}_{i,j}|C_1) &= 1-\theta - P(\mathcal{D}_{i,j}|C_2)
\end{aligned}$$

- All parameters estimated through cross-validation.

Experimental Observations : Competing Methods

- Competing Methods:
  - **Previous Day** as prediction
  - **ARWeather** : Extension of Piorno et al.'s framework to increase prediction horizon to a day
    - Weather attributes with the exception of sunrise and sunset clustered into *NC* groups using *k-means* algorithm
    - Set of weather attributes represented by mean Solar PV value of corresponding cluster
    - To predict multiple hours, use previous prediction as actual value in the next iteration

$$
\begin{aligned}
A_{i,j} \quad = \quad & \beta_1^t * 1 + \beta_2^t * l(i,j) + \beta_3^t * [i - SR(j)] \\
& + \beta_4^t * [ST(j)i] + \beta_5^t * P^a_{i,j,t} + \epsilon
\end{aligned}
$$

Experimental Observations : Competing Methods

- **Stagewise** modeling:
  - Inspired from Stepwise regression . Hocking, 1976
  - Prediction done in correlated stages
  - Stage1: Average Model
  - Stage2: Auto Regression using Solar PV values
  - Stage3: Regression using weather attributes

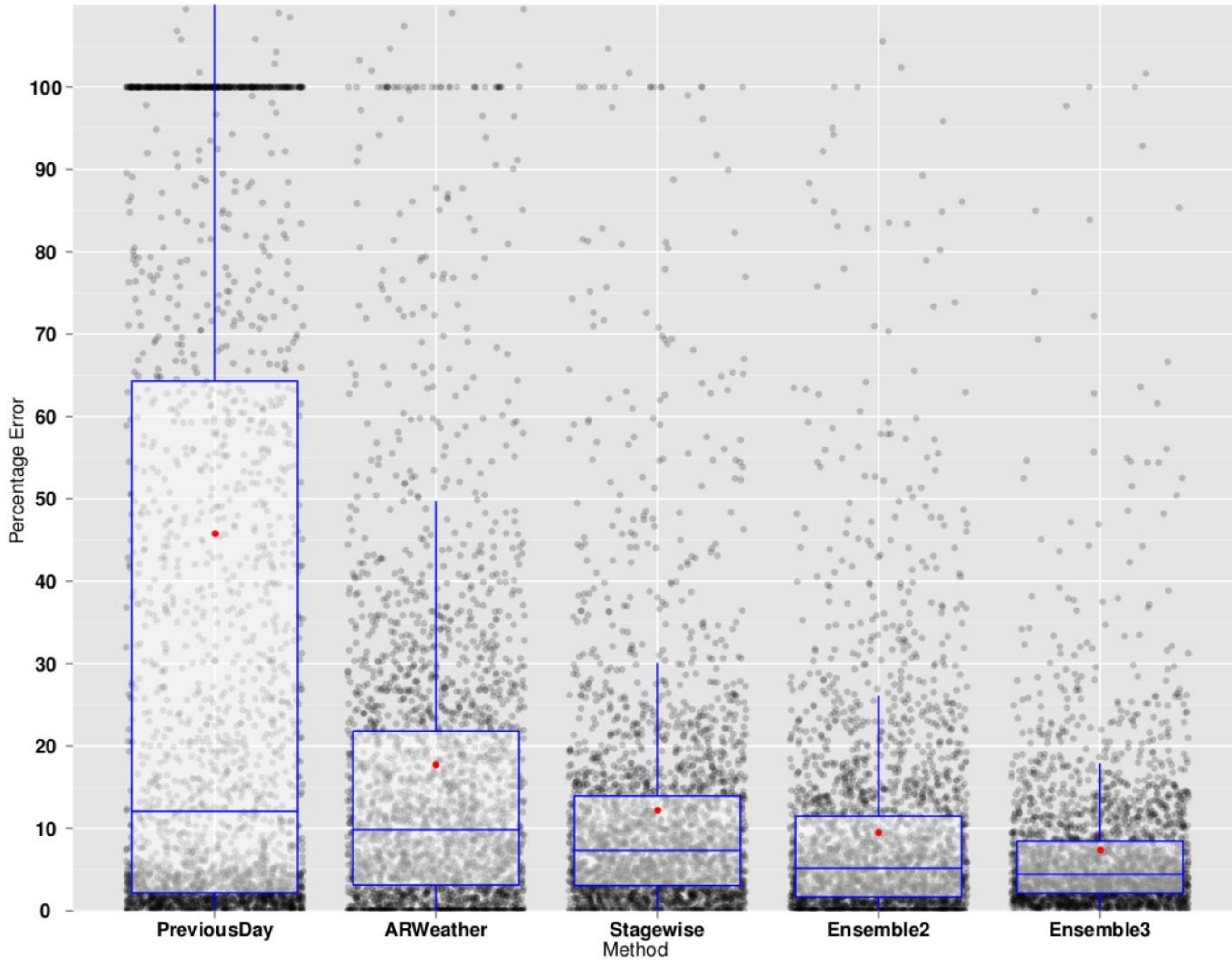
- Bayesian Ensemble methods:
  - **Ensemble2** : only KNN based classifier and Naïve Bayes classifiers used in the ensemble
  - **Ensemble3** : all three classifiers (i.e. including motif based) used in the ensemble

Experimental Observations : Results

Comparative prediction performance at 1-hour offset

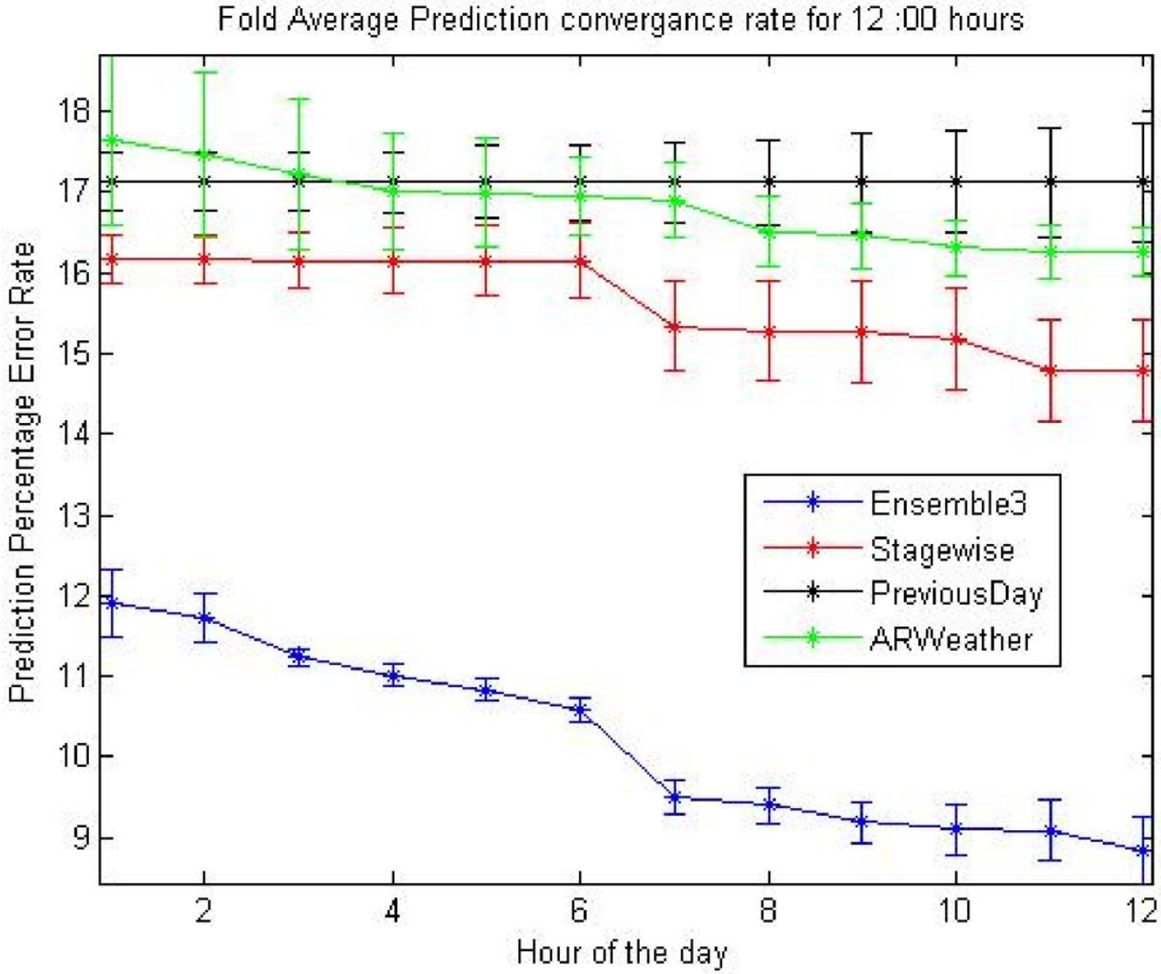
Method	Testing Error		
	Per. Abs. Error	Per. RMS Error	Rel. Abs. Error
PreviousDay	20.54	20.65	20.81
ARWeather	18.54	18.31	19.73
Stagewise	12.77	12.68	15.66
Ensemble2	10.04	10.01	10.01
Ensemble3	8.13	8.21	8.34

Comparative errors of the different methods



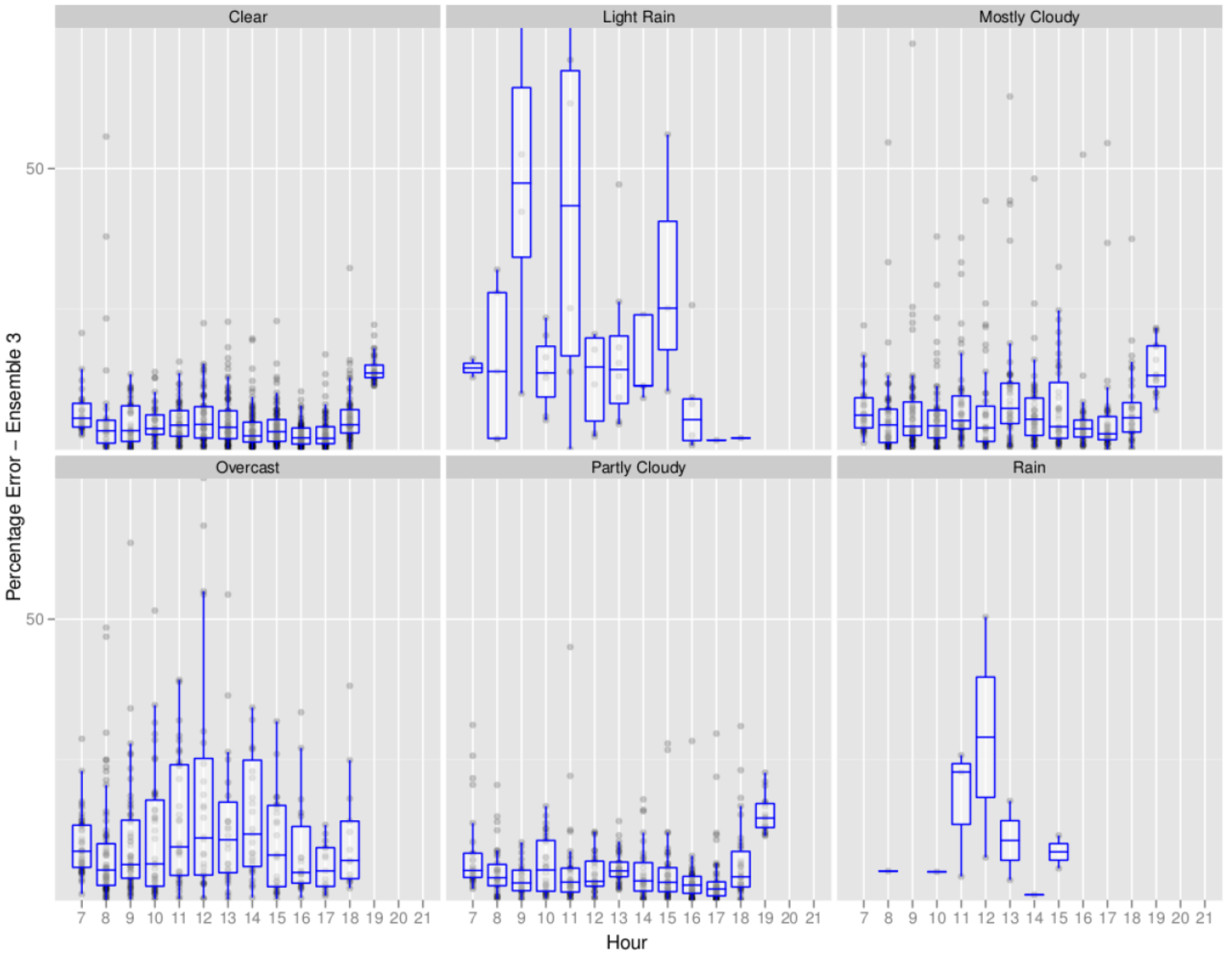
Experimental Observations : Results

Comparative error with different prediction horizon



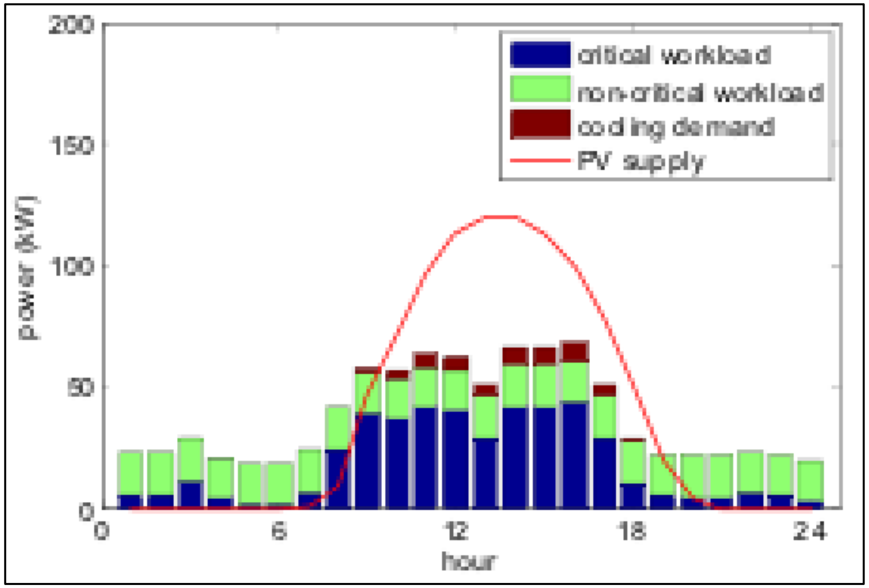
Experimental Observations : Results

Ensemble3: Performance conditioned on weather and hour.

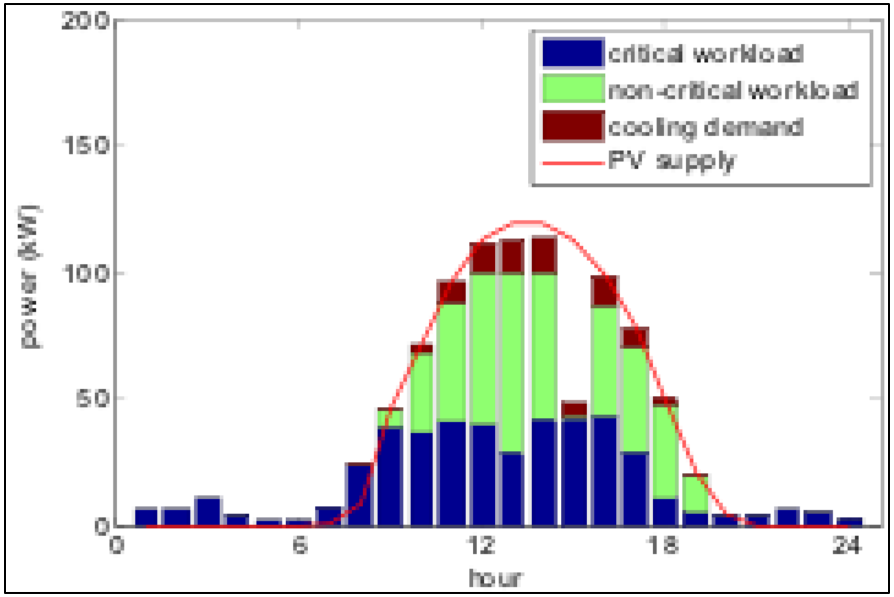


Experimental Observations : Results

Ensemble3: Workload scheduling

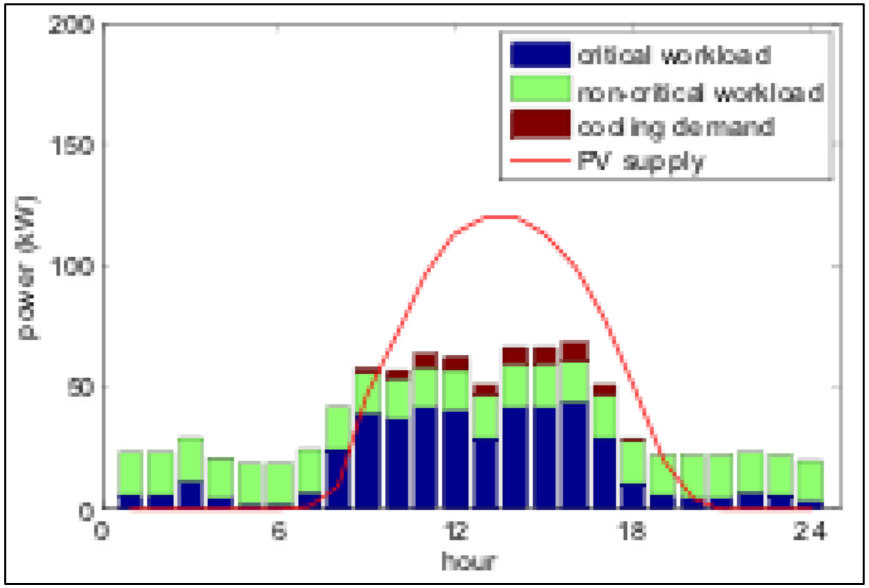


Workload Schedule

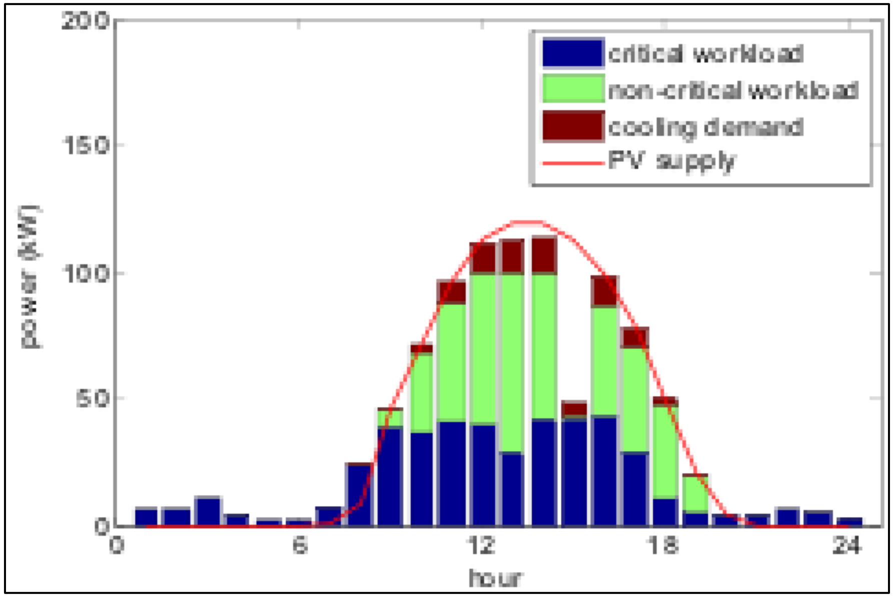


Optimized workload Schedule

Ensemble3: Workload scheduling



Workload Schedule



Optimized workload Schedule

Discussions

- A systematic approach toward Solar PV prediction
- Naturally suited to day-long predictions.
- Three classifiers capture three different nature in the Data
- Ensemble method can incorporate more predictors to improve accuracy

Future Works

- Thoroughly investigate performance of different combinations of the classifiers with respect to different hour and conditions

- Investigate better heuristic to calculate the likelihood of the classifiers for a given day
  - An hour wise confusion matrix for days

- Apply the method on more installations to further analyze the process.

Thank You