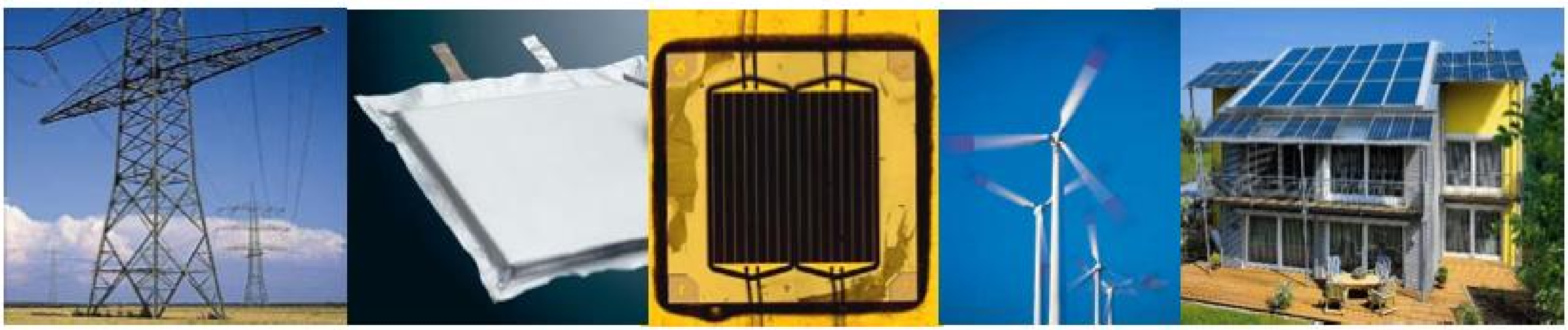
# I know what you did last summer (wanted to save energy): smart meter data accurately predict household intention to enroll in energy efficiency program



*Michael Zeifman*

*Fraunhofer Center for Sustainable Energy Systems*

*BECC 2014*



Fraunhofer USA

# Motivation

- Behavior-based opt-in energy efficiency/demand response programs
  - Evaluation: need for control group
  - Enrollment:
    - High customer acquisition costs
    - Low enrollment rate
- Would love to predict enrollment outcome!

# State of the Art

- Assume the propensity to enroll correlates with household characteristics, e.g., age, household income, education, “environmental issues” (expressed interest in environmental or wildlife issues), “green living” (model-based variable which aims to predict households that are living environmentally friendly), …
- Devise a math model (e.g., logistic regression) to connect these characteristics to probability to enroll
- “Train” the model, i.e., estimate model parameters using household and enrollment data
- Source: M. Harding & A. Hsiaw, “Goal setting and energy conservation.” Journal of Economic Behavior & Organization, in press (2014).

- Problems
  - Prediction accuracy with *training data*: pseudo R-squared 0.035
  - No *testing*
  - Household data: availability and cost

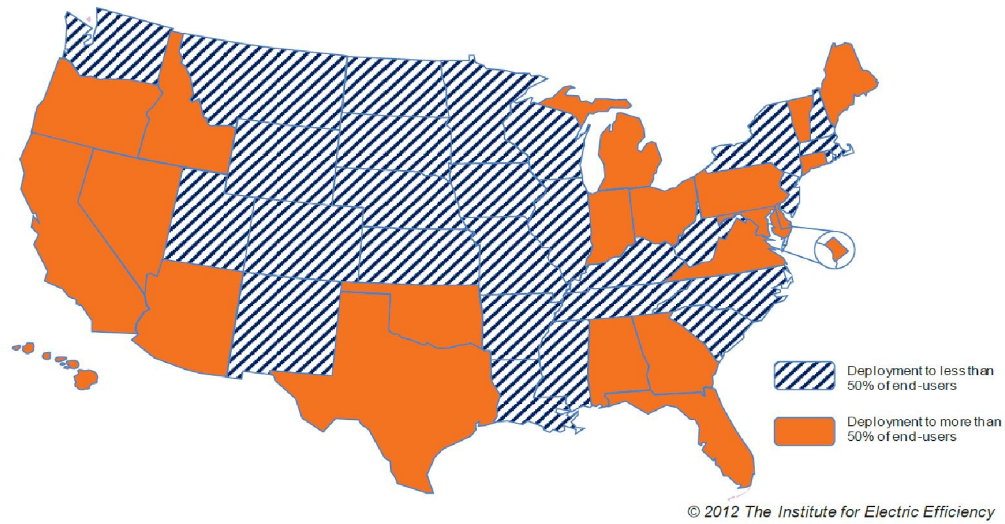
# Objectives

- Accurately predict household propensity to enroll in opt-in behavior based energy efficiency/demand response programs
- Use only data freely available to utilities/program contractors
- Validate method using field data

# The Freely Available Data

Expected Smart Meter Deployments by State by 2015

Source: IEE Report (2012)




- 38,524,639 residential smart meters in the US as of 2012 ([www.eia.gov](http://www.eia.gov))
- Rich interval data embedding household appliance usage behaviors



Fraunhofer USA

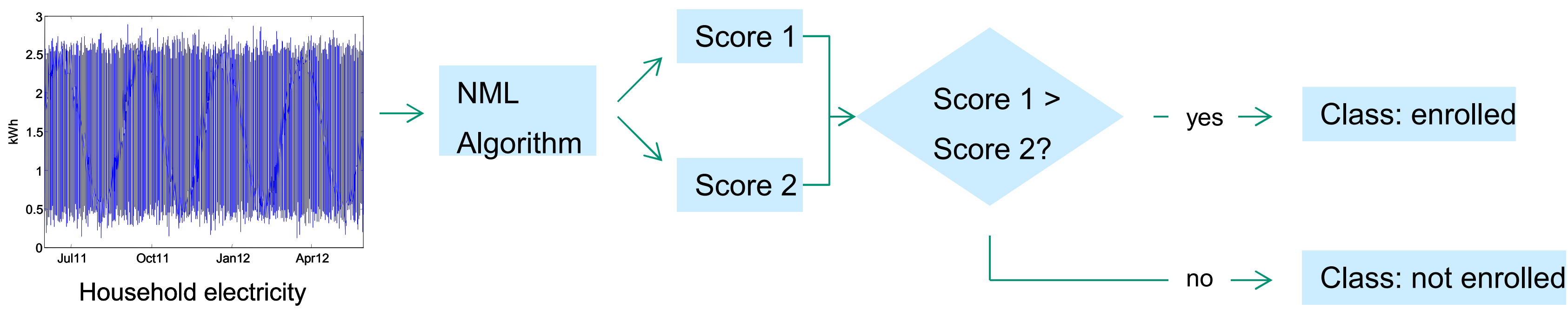
# Proposed Method (1/2)

- State-of-the-art technique (logistic regression) unsuitable
  - Problems with input data (365*24 = 8,760 data points per household)
  - Not optimal for discrimination
- Nonlinear machine learning algorithm*
  - Can handle large data arrays for input
  - Specifically designed for discrimination between 2 classes (enrolled/ not enrolled)
  - Uses training, similarly to logistic regression

*Patent pending

# Proposed Method (2/2)

- Black box system: input –> NML algorithm –> output
  - Input: 1 hour-resolution electricity consumption from a smart meter collected over ~ 1 year
  - Output: Two scores for either class (enrolled/ not enrolled)



Household electricity consumption data

# Case Study: Description

- Fraunhofer CSE was contracted to evaluate energy saving of a new opt-in residential behavior Program.
- Program elements:
  - West-coast based
  - Several recruitment channels – local educational institutes, social media and news advertisement
  - Eligibility: must reside in certain area of a major city
  - Participants control their electricity usage by monitoring their hourly electricity consumption data
  - Significant awards for energy savings to the participating households
- No experimental control group

# Field Data Description

- Pool 1: Smart meter data on ~5,600 households that enrolled in the Program (out of 470,000 eligible households, or 1.2%)
  - Hourly electricity consumption for ~18 months (including 1 year before the Program) of each household
  - Zip code of each household
- Pool 2: Smart meter data on ~32,000 households resided just outside the eligible area (still same city and microclimate zone)
  - Hourly electricity consumption for ~18 months (1 year before the Program) of each household
  - Zip code of each household
- Pool 1 seems to be similar to Pool 2
  - Socio-economic data do not differ significantly (US Census by zip code)
  - Average hourly electricity consumptions do not differ significantly

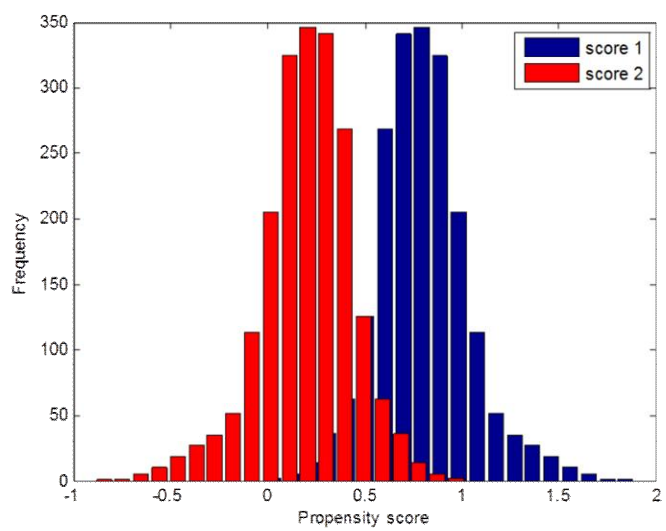
# Algorithm Training

1-year smart meter data

Enrollment matrix

$$\begin{pmatrix} 1\times 8{,}760 \ \textit{data array for enrolled household \#1} \\ \dots\dots\dots\dots \\ 1\times 8{,}760 \ \textit{data array for enrolled household \#2000} \\ \\ 1\times 8{,}760 \ \textit{data array for not enrolled household \#1} \\ \dots\dots\dots\dots \\ 1\times 8{,}760 \ \textit{data array for not enrolled household \#2000} \end{pmatrix} \rightarrow \text{Algorithm} \rightarrow \begin{pmatrix} 1 & 0 \\ \dots & \dots \\ 1 & 0 \\ 0 & 1 \\ \dots & \dots \\ 0 & 1 \end{pmatrix}$$

Results: Distributions of scores (1 - for enrollment class, 2 for non-enrollment class) of trained algorithm for a sample of 2,000 enrolled households.






Fraunhofer USA

# Algorithm testing

- Classify a random sample of 2,000 enrolled households and a random sample of 2,000 not enrolled households that were not used for training (testing samples)
- Repeat the process of training and testing using random samples (multiple cross-validation)

Classification Accuracy in Cross Validation, 95% Confidence Interval

| Samples used | Enrolled households | Not enrolled households |
|---|---|---|
| Training samples | 92.4±1.1 % | 91.7±1.3 % |
| Testing samples | 91.2±1.1 % | 90.5±1.4 % |

## Objectives

- ✓ Accurately predict household propensity to enroll in opt-in energy efficiency/demand response programs
- ✓ Use only data freely available to utilities/program contractors
- ✓ Validate method using field data

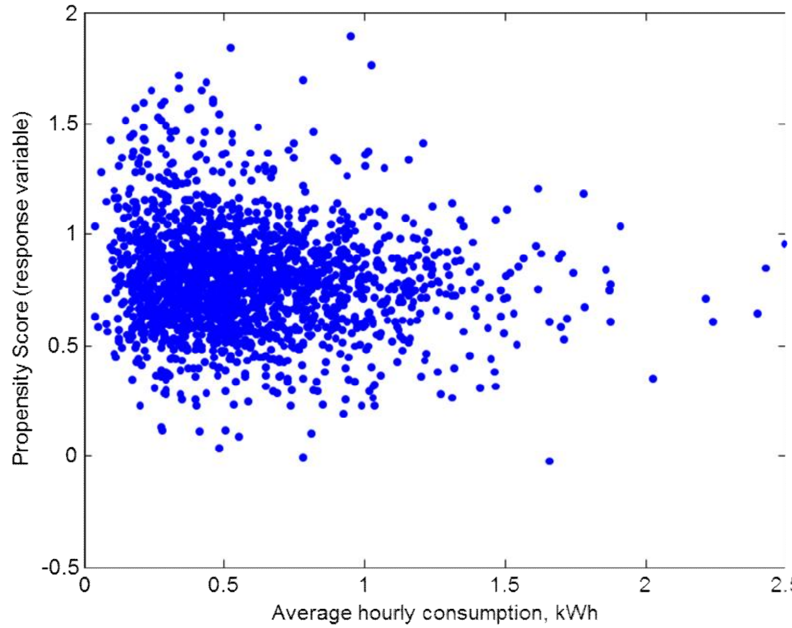
## Potential Implementation

- Sample of ~1,000 households that enrolled – obtain interval pre-program data for ~ 1 year
- Sample of ~1,000 households that did not enroll – obtain interval pre-program data for ~ 1 year
- Train the algorithm ("push button")
- Obtain pool of candidate households (interval data for each household)
- Calculate which households are likely to enroll ("push button")

# Interesting Observations

- Possible correlation between propensity to enroll and average electricity consumption?

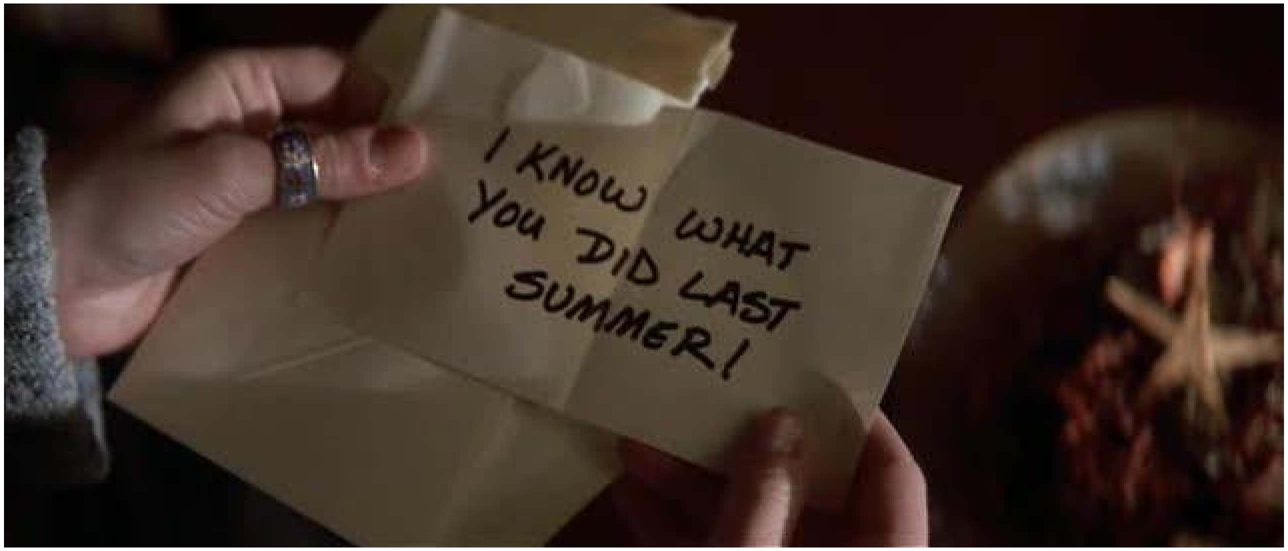
Average hourly electricity consumption vs. propensity score for a sample of 2,000 enrolled households



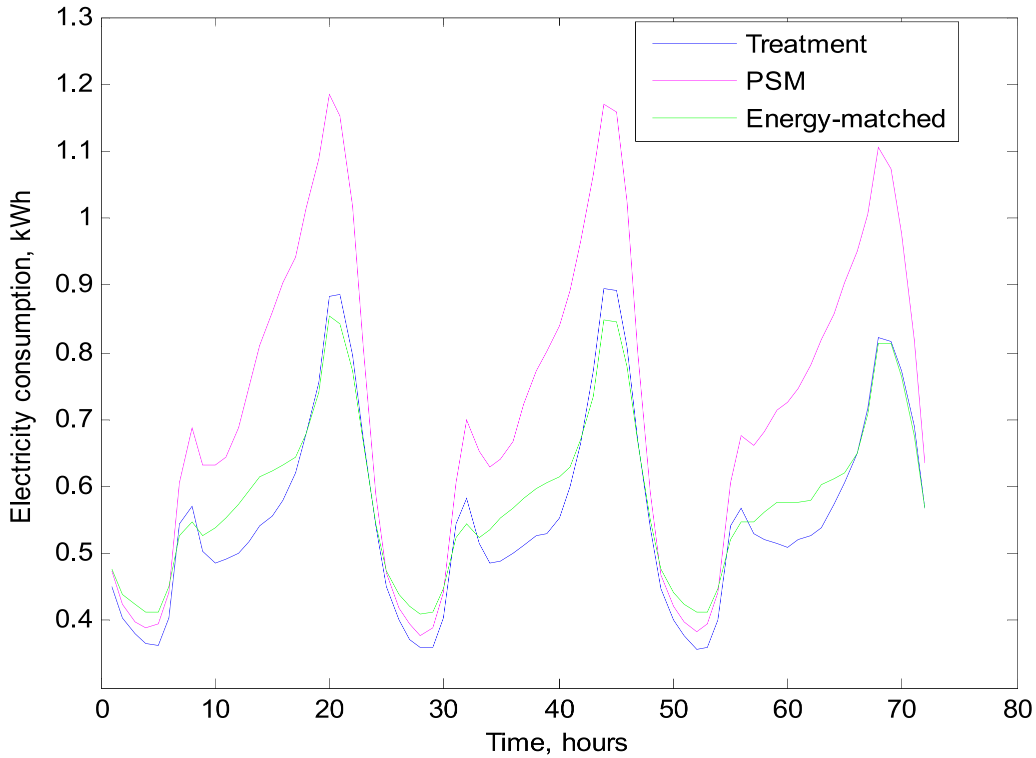

- ~9% are likely to enroll but actual enrollment rate 1.2% ??

# Open Questions

- Our method worked well (enrollment prediction with 90% accuracy) for given region/program. Will it work elsewhere?
- Can enrollment data from one program be used as proxy for another program in the same region?
- Can NML algorithm, trained on data from one region, be used for another region?
- What are requirements to the region (e.g., size, homogeneity)?
- Targeted advertisement for those likely to enroll?


I KNOW WHAT
YOU DID LAST
SUMMER!

# Additional materials




Hourly electricity consumption data for September 21-23, 2011, averaged over groups

- PSM = non-experimental control group built using propensity score from candidate households
- Energy-matched = non-experimental control group build by matching average energy between treatment and candidate households



Fraunhofer USA