A Bayesian Perspective on Residential Demand Response Using Smart Meter Data

Datong-Paul Zhou, Maximilian Balandat, and Claire Tomlin

University of California, Berkeley

[datong.zhou, balandat, tomlin]@eecs.berkeley.edu

September 30, 2016

Residential Demand Response

Background

- Reliable grid operation is dependent on adequate supply and flexible resources

Residential Demand Response

Background

- Reliable grid operation is dependent on adequate supply and flexible resources
- 2009: CAISO proposes "Proxy Demand Resource Product" to "facilitate the participation of existing retail demand programs in the ISO market"
- Proxy Demand Resources (PDRs) offer bids to reflect flexibility to adjust load in response to market schedules

Residential Demand Response

Background

- Reliable grid operation is dependent on adequate supply and flexible resources
- 2009: CAISO proposes "Proxy Demand Resource Product" to "facilitate the participation of existing retail demand programs in the ISO market"
- Proxy Demand Resources (PDRs) offer bids to reflect flexibility to adjust load in response to market schedules
- CPUC Resolution E-4728 (July 2015) approves "an auction mechanism for demand response capacity, called the demand response auction mechanism (DRAM)"
- (Residential) Demand Response Providers with the ability to aggregate customers capable of reducing load participate in the ISO day-ahead, real-time ancillary services market

Residential Demand Response

Background

- Reliable grid operation is dependent on adequate supply and flexible resources
- 2009: CAISO proposes "Proxy Demand Resource Product" to "facilitate the participation of existing retail demand programs in the ISO market"
- Proxy Demand Resources (PDRs) offer bids to reflect flexibility to adjust load in response to market schedules
- CPUC Resolution E-4728 (July 2015) approves "an auction mechanism for demand response capacity, called the demand response auction mechanism (DRAM)"
- (Residential) Demand Response Providers with the ability to aggregate customers capable of reducing load participate in the ISO day-ahead, real-time ancillary services market
- Utilities can obtain resource adequacy through residential resources
- Benchmarking of reduction potential by using consumption "baselines"

Measuring Reduction in Consumption

- Estimate the demand reduction by using suitable baselines (counterfactuals)

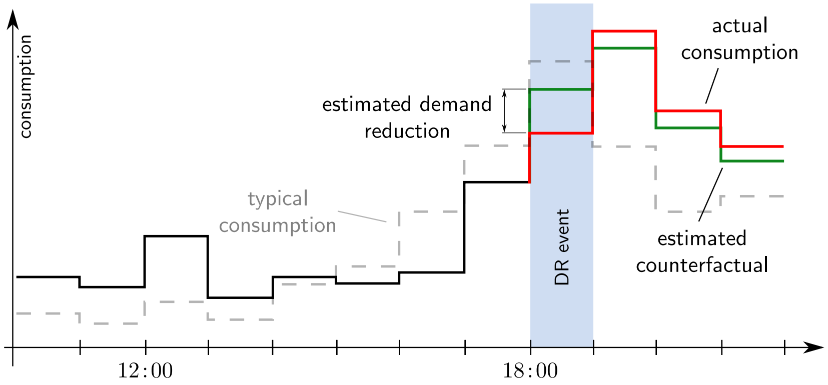
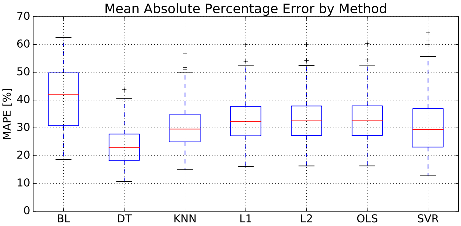


Figure: Courtesy of Maximilian Balandat

Previous Work[1]

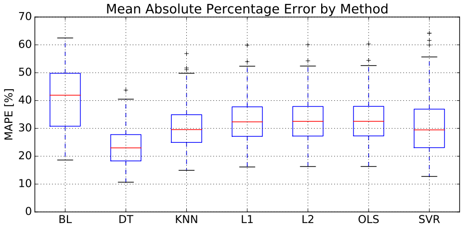
- Short-term load forecasting on the individual household level



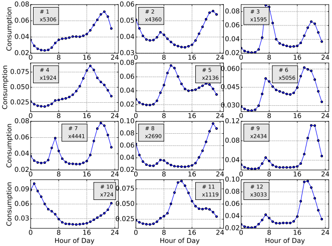
[1] D. Zhou, M. Balandat, and C. Tomlin. **Residential Demand Response Targeting Using Observational Data, 55th Conference on Decision and Control, December 2016** (to appear).

Previous Work[1]

- Short-term load forecasting on the individual household level



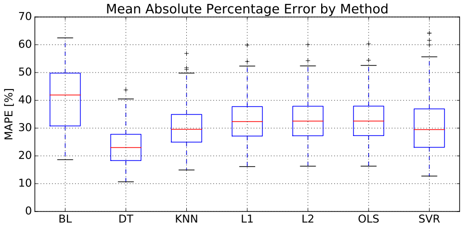
- K-means clustering of daily electricity load shapes



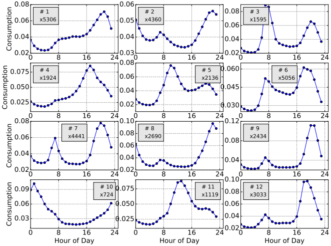
[1] D. Zhou, M. Balandat, and C. Tomlin. **Residential Demand Response Targeting Using Observational Data, 55th Conference on Decision and Control, December 2016** (to appear).

Previous Work[1]

- Short-term load forecasting on the individual household level



- *K*-means clustering of daily electricity load shapes



- Estimation of counterfactual consumption to estimate DR reduction
- Measured the variability of users with entropy
- *Targeting*: Users with high variability seem to reduce more during DR events

[1] D. Zhou, M. Balandat, and C. Tomlin. **Residential Demand Response Targeting Using Observational Data, 55th Conference on Decision and Control, December 2016** (to appear).

Today's Talk

Agenda

- Overview of Machine Learning tools for short-term load forecasting

Today's Talk

Agenda

- Overview of Machine Learning tools for short-term load forecasting
- Inclusion of physically motivated latent variables to improve prediction accuracy
- Key idea: Estimate whether a user has "high" or "low " consumption
  - If consumption at t is high, it is probably high at $t+1$, as well
  - If consumption at t is low, it is probably low at $t+1$, as well

Today's Talk

Agenda

- Overview of Machine Learning tools for short-term load forecasting
- Inclusion of physically motivated latent variables to improve prediction accuracy
- Key idea: Estimate whether a user has "high" or "low " consumption
  - If consumption at t is high, it is probably high at $t+1$, as well
  - If consumption at t is low, it is probably low at $t+1$, as well
- Bayesian Perspective in STLF: Encode this consumption variable as an unobserved, latent variable
  - Conditional Gaussian Mixture Model
  - Hidden Markov Model

Today's Talk

Agenda

- Overview of Machine Learning tools for short-term load forecasting
- Inclusion of physically motivated latent variables to improve prediction accuracy
- Key idea: Estimate whether a user has "high" or "low " consumption
  - If consumption at t is high, it is probably high at $t+1$, as well
  - If consumption at t is low, it is probably low at $t+1$, as well
- Bayesian Perspective in STLF: Encode this consumption variable as an unobserved, latent variable
  - Conditional Gaussian Mixture Model
  - Hidden Markov Model
- Forecasting Algorithms to estimate DR Reduction
- Numerical Results of DR Case Study

Machine Learning for Short-Term Load Forecasting

- **Regression Problem**: $y = f(X, \beta)$, where X is the covariate matrix

Machine Learning for Short-Term Load Forecasting

- **Regression Problem**: $y = f(X, \beta)$, where X is the covariate matrix
- **Ordinary Least Squares**

$$\hat{\beta}_{OLS} = \arg\min_{\beta} \|y - X\beta\|_2^2$$

- **Ridge (L2) and Lasso (L1) Regression**

$$\hat{\beta}_{\text{L1}} = \arg\min_{y} \|y - X\beta\|_2^2 + \lambda\|\beta\|_1 \rightarrow \hat{y}_{\text{L1}} = X\hat{\beta}_{\text{L1}}$$

$$\hat{\beta}_{\text{L1}} = \arg\min_{y} \|y - X\beta\|_2^2 + \lambda\|\beta\|_2 \rightarrow \hat{y}_{\text{L2}} = X\hat{\beta}_{\text{L2}}$$

Machine Learning for Short-Term Load Forecasting

- **Regression Problem**: $y = f(X, \beta)$, where X is the covariate matrix
- **Ordinary Least Squares**

$$\hat{\beta}_{OLS} = \arg\min_{\beta} \|y - X\beta\|_2^2$$

- **Ridge (L2) and Lasso (L1) Regression**

$$\hat{\beta}_{\text{L1}} = \arg\min_{y} \|y - X\beta\|_2^2 + \lambda\|\beta\|_1 \rightarrow \hat{y}_{\text{L1}} = X\hat{\beta}_{\text{L1}}$$
$$\hat{\beta}_{\text{L1}} = \arg\min_{y} \|y - X\beta\|_2^2 + \lambda\|\beta\|_2 \rightarrow \hat{y}_{\text{L2}} = X\hat{\beta}_{\text{L2}}$$

- k-**Nearest Neighbors Regression**

$$\hat{y}_{\text{KNN}} = (y_1 + \ldots + y_k)/k$$

- **Decision Tree Regression**: Cross-validation on max. depth and min. samples per node
- **Support Vector Regression**: Cross-validation on slack variables

Latent Variable Models - Mixtures

Mixture of Linear Regression Models

- Consider K linear regression models, each governed by own weight w_k
- Assume a *common* noise variance σ^2
- Mixture distribution with mixing proportions $\{\pi_k\}$:

$$\mathbb{P}(y|\mathbf{w}, \sigma^2, \pi) = \sum_{k=1}^{K} \pi_k \mathcal{N}(y|w_k \cdot x, \sigma^2)$$

Latent Variable Models - Mixtures

Mixture of Linear Regression Models

- Consider K linear regression models, each governed by own weight w_k
- Assume a *common* noise variance σ^2
- Mixture distribution with mixing proportions $\{\pi_k\}$:

$$\mathbb{P}(y|\mathbf{w}, \sigma^2, \pi) = \sum_{k=1}^{K} \pi_k \mathcal{N}(y|w_k \cdot x, \sigma^2)$$

- Idea: Interpret electricity consumption as a result of behavioral archetypes $k = 1, \ldots, K$
- Expectation-Maximization Algorithm yields update rules:

$$\gamma_{nk} := \mathbb{E}[z_{nk}] = \mathbb{E}\left[\ell(z|x_n, \theta^{\text{old}})\right] = \frac{\pi_k \mathcal{N}(y_n|w_k \cdot x_n, \sigma^2)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(y_n|w_j \cdot x_n, \sigma^2)}$$

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} \gamma_{nk}, \quad w_k = \left[X^\top D X\right]^{-1} X^\top D Y, \quad D = \text{diag}(\gamma_{1k}, \ldots, \gamma_{nk})$$

$$\sigma^2 = \frac{1}{N} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma_{nk} (y_n - w_k \cdot x_n)^2$$

Latent Variable Models - Mixtures (cont'd.)

Mixture Models for Prediction

- IID assumption of Mixture Models $\Rightarrow$ cannot capture the temporal correlation in time series data
- Prediction Method 1: Convex combination of learners:

$$\hat{y} = \sum_{k=1}^{K} \hat{\pi}_k \hat{w}_k \cdot x$$

Latent Variable Models - Mixtures (cont'd.)

Mixture Models for Prediction

- IID assumption of Mixture Models $\Rightarrow$ cannot capture the temporal correlation in time series data
- Prediction Method 1: Convex combination of learners:

$$\hat{y} = \sum_{k=1}^{K} \hat{\pi}_k \hat{w}_k \cdot x$$

- Prediction Method 2: If model i's covariates are "spatially separated" from model j:

$$j = \arg\min_{1 \leq i \leq N} \|x_i - x\|_2$$

$$\hat{y} = \sum_{k=1}^{K} \gamma_{jk} \hat{w}_k \cdot x$$

Latent Variable Models - Hidden Markov Models

- Data often is not i.i.d.
- Use on sequential data, e.g. time series, speech recognition, DNA sequences

Latent Variable Models - Hidden Markov Models

- Data often is not i.i.d.
- Use on sequential data, e.g. time series, speech recognition, DNA sequences

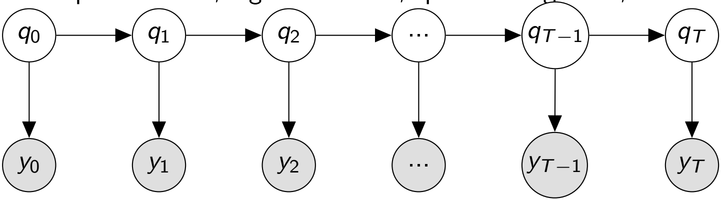


Figure: Hidden Markov Model. Hidden States q, Observations y

Latent Variable Models - Hidden Markov Models

- Data often is not i.i.d.
- Use on sequential data, e.g. time series, speech recognition, DNA sequences

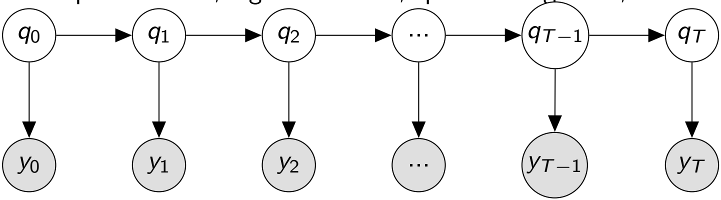


Figure: Hidden Markov Model. Hidden States q, Observations y

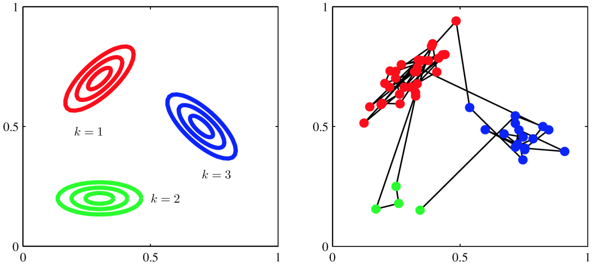


Figure: 3 state latent variable, Gaussian emission model

Latent Variable Models - Hidden Markov Models (cont'd.)

Elements of HMMs

- Hidden Layer: Transition between states is governed by a Markov Process

$$a_{ij} = \mathbb{P}(q_t = j | q_{t-1} = i),\ a_{ij} > 0, \quad 1 \leq i, j \leq M,\ t = 0, 1, 2, \ldots, N$$

Latent Variable Models - Hidden Markov Models (cont'd.)

Elements of HMMs

- Hidden Layer: Transition between states is governed by a Markov Process

$$a_{ij} = \mathbb{P}(q_t = j | q_{t-1} = i), \; a_{ij} > 0, \quad 1 \leq i, j \leq M, \; t = 0, 1, 2, \ldots, N$$

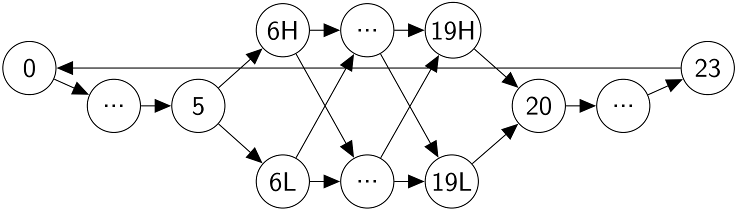


Figure: Markov Transition Diagram, 24 hour periodicity

Latent Variable Models - Hidden Markov Models (cont'd.)

Elements of HMMs

- Hidden Layer: Transition between states is governed by a Markov Process

$$a_{ij} = \mathbb{P}(q_t = j | q_{t-1} = i), \ a_{ij} > 0, \quad 1 \leq i, j \leq M, \ t = 0, 1, 2, \ldots, N$$

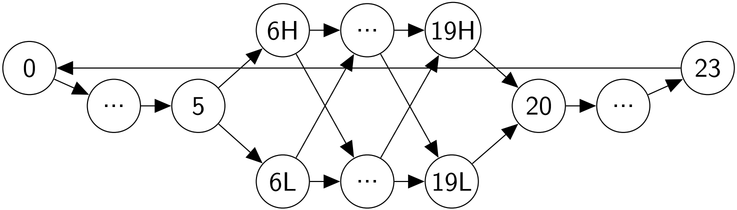


Figure: Markov Transition Diagram, 24 hour periodicity

- Observables: Hidden states "emit" observations according to some distribution, e.g. Gaussian:

$$\mathbb{P}(y_t = y | q_t = q) = \frac{1}{\sigma_q \sqrt{2\pi}} \exp\left(-\frac{(y - \mu_q)^2}{2\sigma_q^2}\right)$$

Latent Variable Models - Hidden Markov Models (cont'd.)

Parameter Estimation

- Maximum Likelihood Estimation and EM Algorithm
- Use Bayes Rule and conditional independencies of graphical model:

$$\mathbb{P}(q_t|y) =: \frac{\alpha(q_t)\mathbb{P}(y_t|q_t)\beta(q_t)}{\mathbb{P}(y)}$$

$$\mathbb{P}(q_t, q_{t+1}|y) = \frac{\alpha(q_t)\beta(q_{t+1})a_{q_t,q_{t+1}}\mathbb{P}(y_t|q_t)\mathbb{P}(y_{t+1}|q_{t+1})}{\mathbb{P}(y)}$$

- Compute $\alpha(q_t), \beta(q_t)$ with the famous α-β-recursion

Latent Variable Models - Hidden Markov Models (cont'd.)

Parameter Estimation

- Maximum Likelihood Estimation and EM Algorithm
- Use Bayes Rule and conditional independencies of graphical model:

$$\mathbb{P}(q_t|y) =: \frac{\alpha(q_t)\mathbb{P}(y_t|q_t)\beta(q_t)}{\mathbb{P}(y)}$$

$$\mathbb{P}(q_t, q_{t+1}|y) = \frac{\alpha(q_t)\beta(q_{t+1})a_{q_t,q_{t+1}}\mathbb{P}(y_t|q_t)\mathbb{P}(y_{t+1}|q_{t+1})}{\mathbb{P}(y)}$$

- Compute $\alpha(q_t), \beta(q_t)$ with the famous α-β-recursion

Inference: Filtering, Predicting, Smoothing

$$\mathbb{P}(q_t|y_0, \ldots, y_t) = \frac{\alpha(q_t)\mathbb{P}(y_t|q_t)}{\mathbb{P}(y_0, \ldots, y_t)}$$

$$\mathbb{P}(q_{t+1}|y_0, \ldots, y_t) = \frac{\alpha(q_{t+1})}{\mathbb{P}(y_0, \ldots, y_t)}$$

$$\mathbb{P}(q_t|y_0, \ldots, y_n) = \frac{\alpha(q_t)\mathbb{P}(y_t|q_t)\beta(q_t)}{\mathbb{P}(y_0, \ldots, y_n)}, \quad 0 < n < t$$

HMM – Case Study

Application of HMM on electricity consumption of residential customers

- Fit a HMM on the electricity consumption of 273 customers
- Use estimated latent variable as an additional covariate for parametric regression on electricity consumption

HMM – Case Study

Application of HMM on electricity consumption of residential customers

- Fit a HMM on the electricity consumption of 273 customers
- Use estimated latent variable as an additional covariate for parametric regression on electricity consumption

Algorithm 1 Real-Time Prediction

procedure FITREGRESSION(T)
  Split observations into training and test sequence: $(y_0, \dots, y_T)$, $(y_{T+1}, \dots, y_N)$
  Fit HMM on training sequence $(y_0, \dots, y_T)$
  Estimate Hidden States $(q_0, \dots, q_T)$
  Model 1 $\leftarrow$ Fit Regression without $(q_0, \dots, q_T)$
  Model 2 $\leftarrow$ Fit Regression with $(q_0, \dots, q_T)$
procedure STEPWISEPREDICTION(U)
  for $i = T + 1, \dots, N$ **do**
    Predict y_i with *Model 1*
    Predict hidden state q_i
    Predict y_i with *Model 2*
    if $i - T \ \% \ U = 0$ **then**
      Retrain HMM on $y_0, \dots, y_i$
      Model 1 $\leftarrow$ Retrain without $(q_0, \dots, q_i)$
      Model 2 $\leftarrow$ Retrain with $(q_0, \dots, q_i)$

HMM – Case Study

Application of HMM on electricity consumption of residential customers

- Fit a HMM on the electricity consumption of 273 customers
- Use estimated latent variable as an additional covariate for parametric regression on electricity consumption

Algorithm 1 Real-Time Prediction

procedure FitRegression(T)
    Split observations into training and test sequence: $(y_0, \ldots, y_T)$, $(y_{T+1}, \ldots, y_N)$
    Fit HMM on training sequence $(y_0, \ldots, y_T)$
    Estimate Hidden States $(q_0, \ldots, q_T)$
    Model 1 ← Fit Regression without $(q_0, \ldots, q_T)$
    Model 2 ← Fit Regression with $(q_0, \ldots, q_T)$
procedure StepwisePrediction(U)
    for $i = T + 1, \ldots, N$ **do**
        Predict y_i with *Model 1*
        Predict hidden state q_i
        Predict y_i with *Model 2*
        if $i - T \ \% \ U = 0$ **then**
            Retrain HMM on $y_0, \ldots, y_i$
            Model 1 ← Retrain without $(q_0, \ldots, q_i)$
            Model 2 ← Retrain with $(q_0, \ldots, q_i)$

Covariates

- 5 previous hourly consumptions
- 5 previous hourly ambient air temperatures
- Model 1: Hour of Day, one-hot encoded
- Model 2: Hour of Day interacted with latent variable, one-hot encoded

HMM – Results of Case Study (cont'd.)

Comparison of Prediction Accuracy

- *Mean Absolute Percentage Error* (MAPE):

$$\mathrm{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{\hat{y}_t - y_t}{y_t}\right|$$

HMM – Results of Case Study (cont'd.)

Comparison of Prediction Accuracy

- *Mean Absolute Percentage Error* (MAPE):

$$\mathrm{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{\hat{y}_t - y_t}{y_t}\right|$$

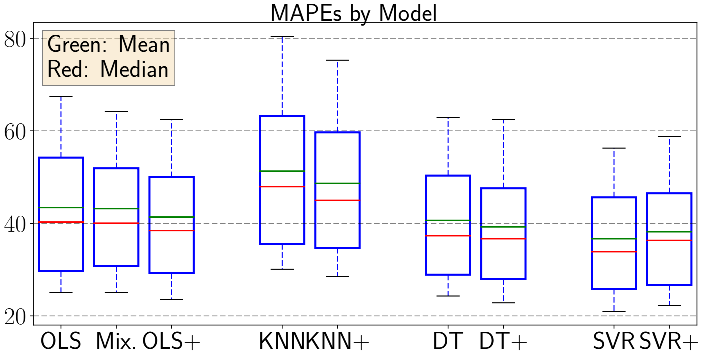


Figure: MAPEs across 273 users for different ML methods and with / without latent variable

HMM – Results of Case Study (cont'd.)

Estimated Reductions During DR Hours by Hour of Day

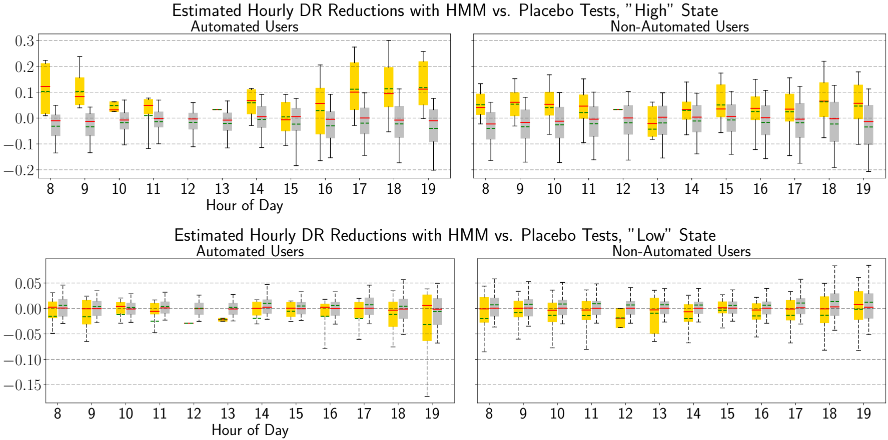


Figure: Boxplots for estimated DR reduction by hour of the day. Top row: "High" consumption, Bottom Row: "Low" consumption, Left Column: Automated users, Right Column: Non-automated users

The End
Questions?