Deep Latent Generative Models For Energy Disaggregation

Gissella Bejarano, David Defazio, Arti Ramesh

BINGHAMTON
UNIVERSITY
STATE UNIVERSITY OF NEW YORK

Overview

- Motivation
- Related Work
- My Contribution
  - Main differences
  - VRNN applied to Generation/Disaggregation
- Results

Energy Disaggregation Problem

- Also called non-intrusive load monitoring or NILM (Hart, 1992)
  - An alternative to submeters
- Disaggregated bills can help **reduce consumption**
  - Home savings
  - Reduce environmental impact

Energy Disaggregation Problem

- Given Aggregated signal

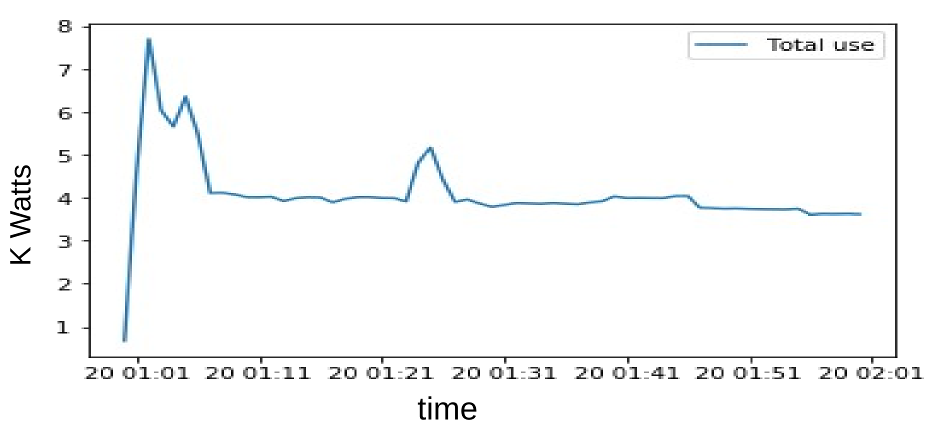
$$\mathbf{x} = (x_1, x_2, ..., x_T) \qquad x_t \in \mathbb{R}_+$$

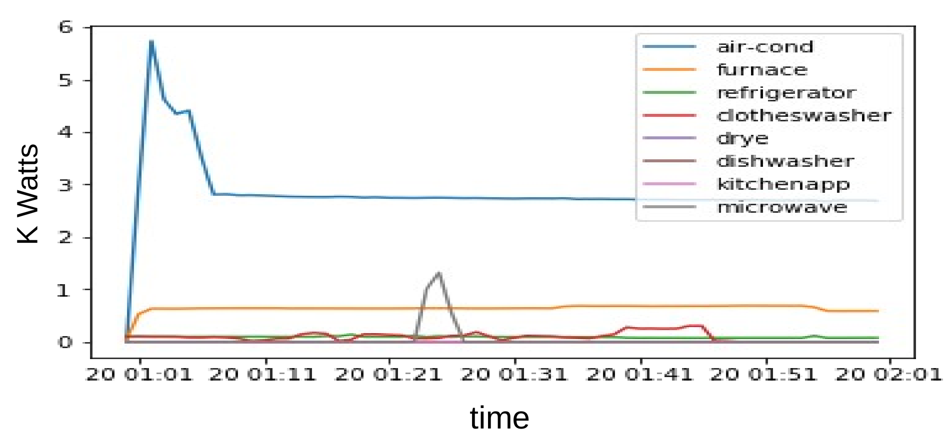
- Infer Disaggregated signal of the appliances

$$\mathbf{y} = (y_1, y_2, ..., y_T) \qquad y_t \in \mathbb{R}_+^I$$

- Relation between x and y

$$x_t = \sum_{i=1}^{I} y_t^{(i)} \qquad I : \#appliances$$



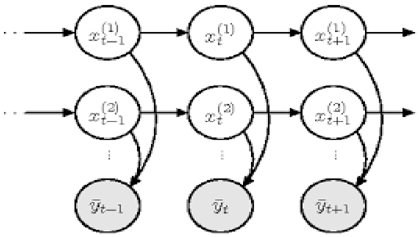


Overview

- Motivation
- Related Work
- My Contribution
  - Main differences
  - VRNN applied to Generation/Disaggregation
- Future Work

Probabilistic and Deep Learning approaches

- Factorial Hidden Markov Model (FFHM)
  - Discrete states (ON/OFF/..)



- Hinge-Loss Markov Random Field (HL-MRF)
  - Discrete states (ON/OFF/..)
  - Discrete intervals
- Latent Bayesian Melding
  - Require 2 datasets (population and individual)

1. Kolter and Jaakkola, 2012, AISTATS. Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation.
2. Tomkins et al., 2017, IJCAI. Disambiguating Energy Disaggregation: A Collective Probabilistic Approach
3. Mingjun Zhong, Nigel Goddard, Charles Sutton, 2015, NIPS. Latent Bayesian melding for integrating individual and population models.

Deep Learning approaches

- Denoising Auto-Encoders (AE)
  - One appliance at a time
  - Deterministic result
- Recurrent Neural Networks (bidirectional LSTM)
  - One appliance at a time
  - Deterministic result

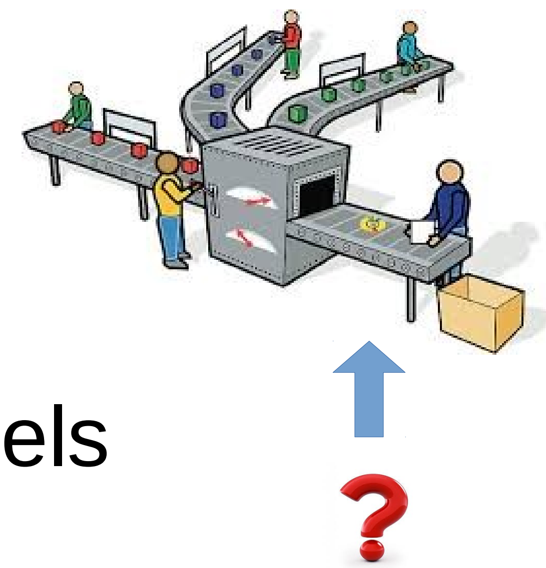
Kelly and Knottenbelt, 2015, BuildSys. Neural NILM: Deep Neural Networks Applied to Energy Disaggregation

Overview

- Motivation
- Related Work
- My Contribution
  - Main differences
  - VRNN applied to Generation/Disaggregation
- Future Work

What do I do differently?

- Do Disaggregation (inference) and generation **jointly**
- Model **continuous values** of all the appliances at the time
- Deep **latent** generative models
  - Stochastic transformation of input
  - Better understanding of the essence
- Deep learning + Probabilistic Models



Contributions

1) Adapt the Variational Recurrent Neural Network model (VRNN) for energy consumption signal generation (i.e. synthetic data)

2) Propose an Energy disaggregation model through a generation model (VRNN)*

- One appliance at the time

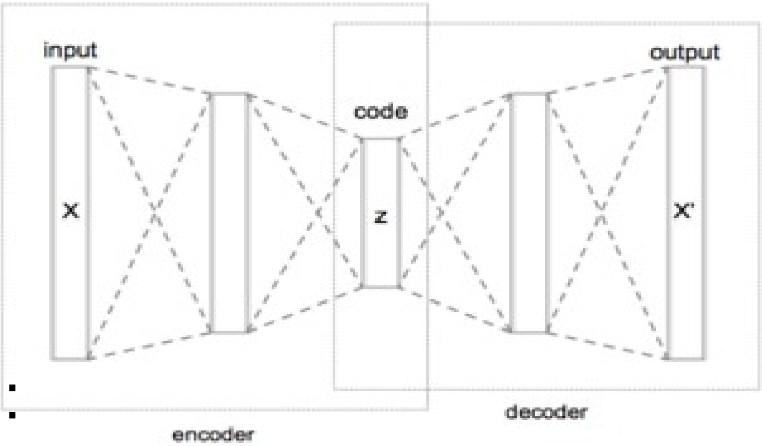
- All appliances at the time

*VRNN only has been used for generation

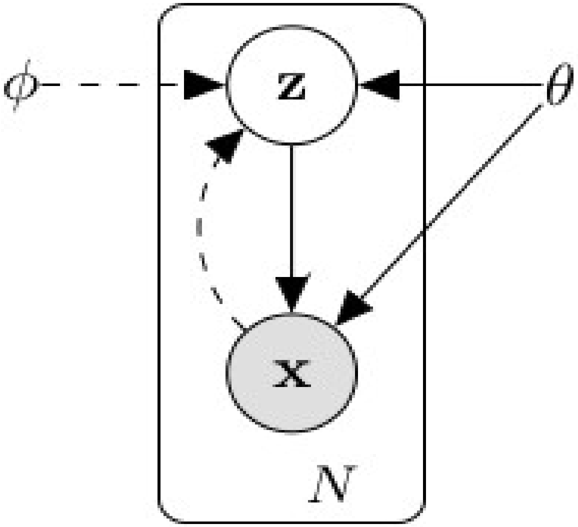
Overview

- Motivation
- Related Work
- My Contribution
  - Main differences
  - VRNN applied to Generation/Disaggregation
- Future Work

VAE



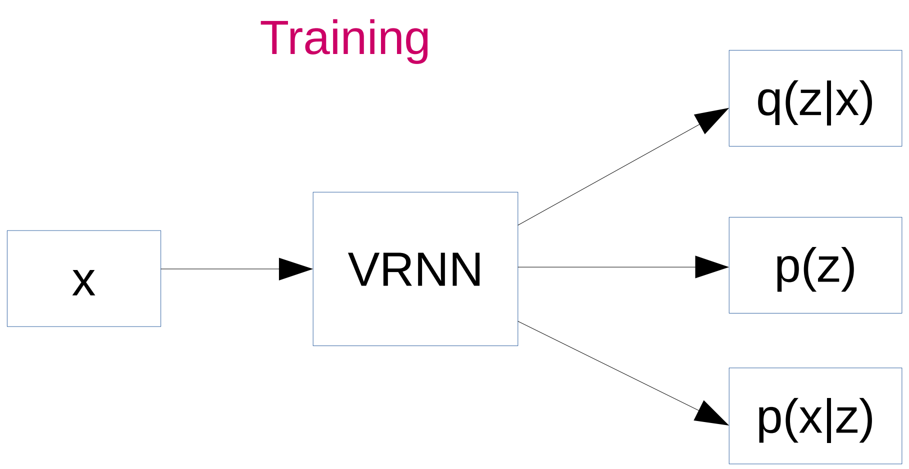
- AE model for dimensionality reduction, denoising: Encoder + Decoder
- VAE Introduces a latent variable to learn:
  - Parameters θ of the generative model $\mathrm{p}_\theta(z)\,\mathrm{p}_\theta(x|z)$
  - Parameters ϕ of the variational approximation $\mathrm{q}_\phi(z|x)$ to the intractable posterior $\mathrm{p}_\theta(z|x)$

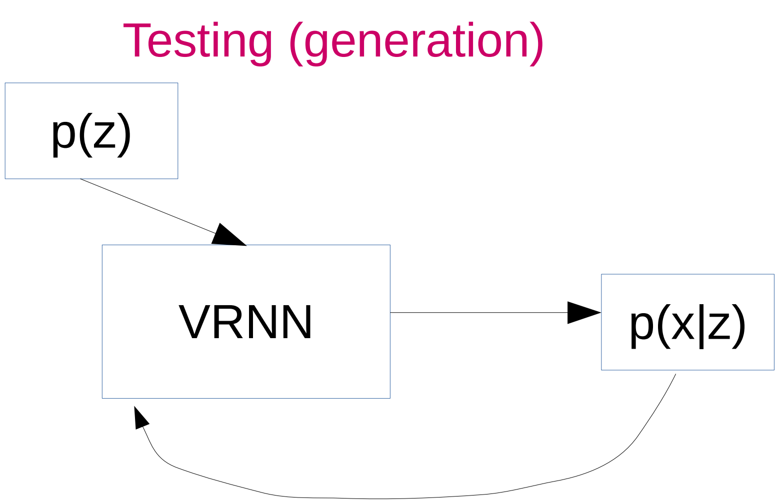


D. P. Kingma and M. Welling, 2014, ICLR. Auto-Encoding Variational Bayes.

VRNN

- It contains a VAE at each time step
- Intermediate layers outputs: Gaussian Mixture Model $$[\mu_{x,t}, \sigma^2_{x,t}] = \sum_{i=1}^{k} \pi_i \phi^{dec}_{i,\tau}$$
- Sample z, x' from those distribution parameters





Cost – VAE vs VRNN

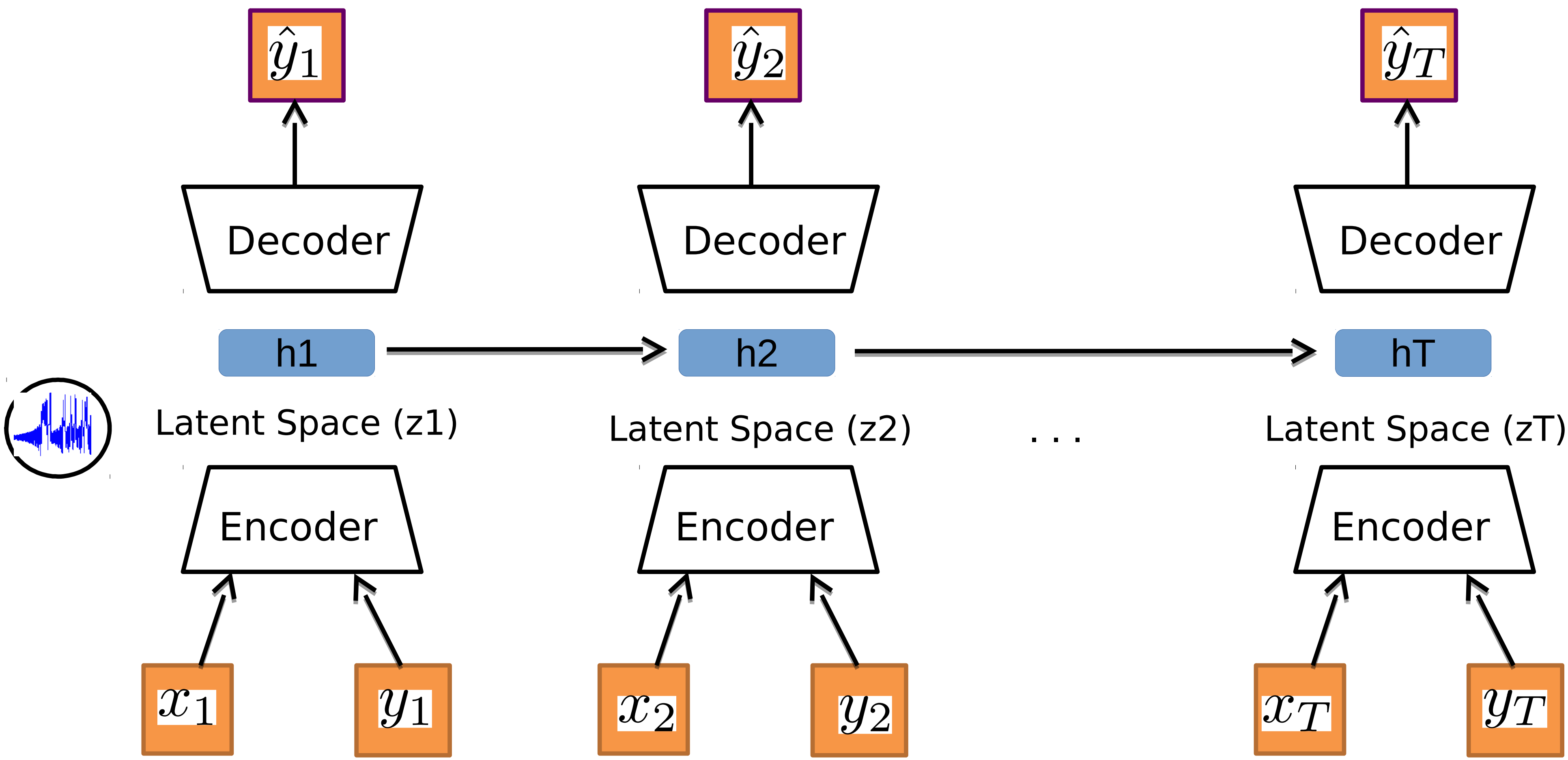
- Maximize Variational Lower bound
  - Kullback–Leibler divergence (q - p) minimization
  - Negative log-likelihood to better explain decoding
- VRNN cost includes previous time-step values

LOG-LIKELIHOOD	
VAE	$\log \mathrm{p}_\theta(x\|z)$
VRNN	$\log \mathrm{p}_\theta(x_t\|z_{\leq t}, x_{<t})$

KL divergence	
VAE	$\mathrm{D}_{KL}(q_\phi(z\|x)\\|p_\theta(z))$
VRNN	$\mathrm{D}_{KL}(q_\phi(z_t\|x_{\leq t}, z_{<t})\\|p_\theta(z_t\|x_{<t}, z_{<t}))$

General VRNN for Disaggregation

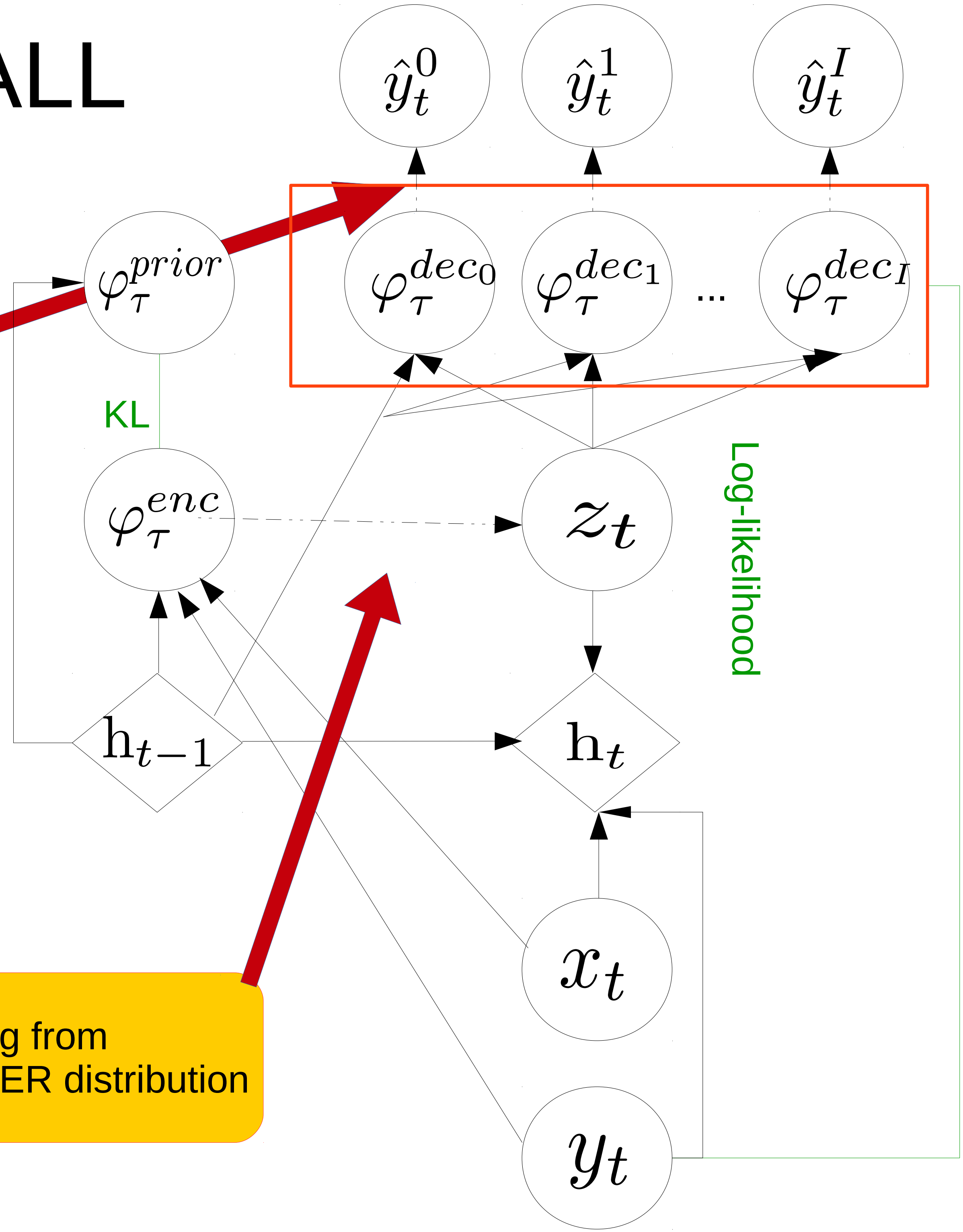
- VAE at each time step



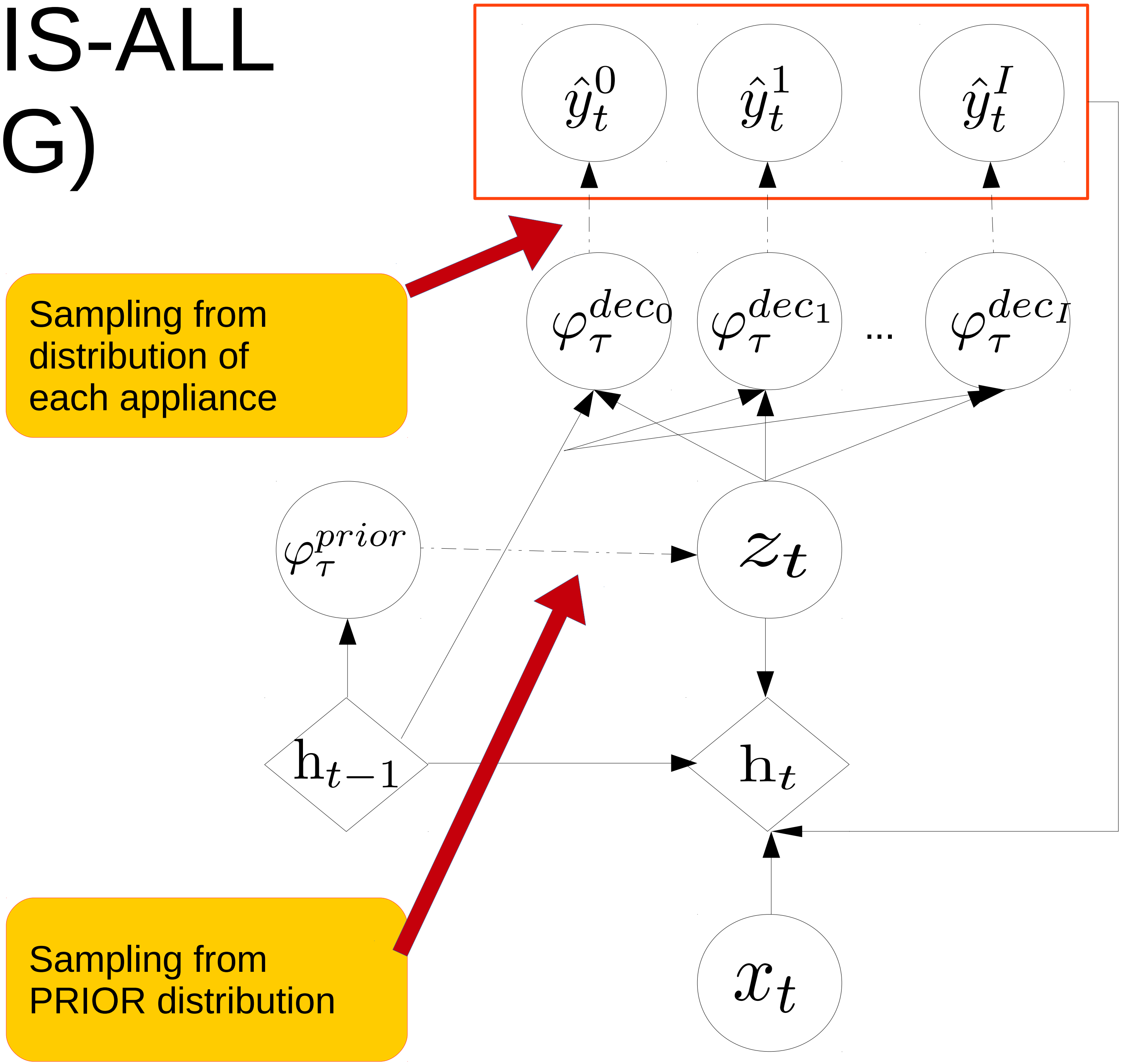
VRNN-DIS-ALL (TRAINING)

Sampling from distribution of each appliance

Sampling from ENCODER distribution



VRNN-DIS-ALL (TESTING)



Overview

- Motivation
- Related Work
- My Contribution
  - Main differences
  - VRNN applied to Generation/Disaggregation
- Results

Datasets

REDD

- 4 buildings
- 4 appliances
- Samples every 1 minute
- ~2 months

Dataport (Pecan-Street)

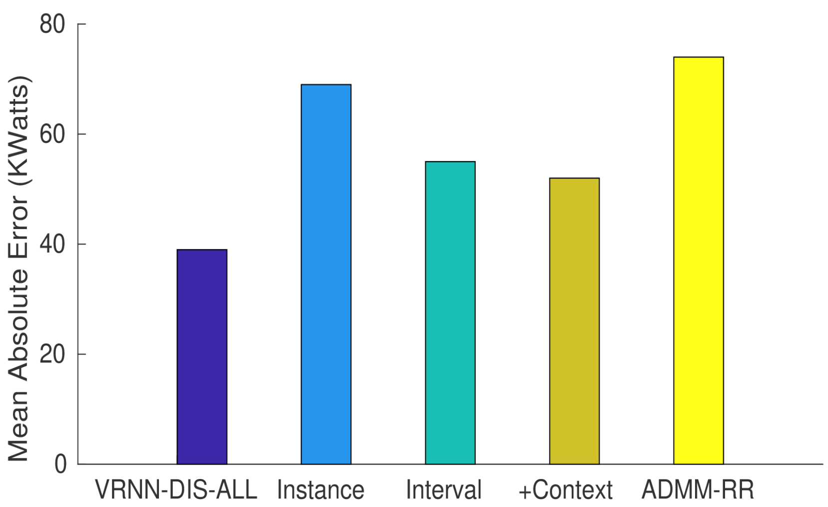
- 5 buildings
- 8 appliances
- Samples every 1 minute
- 1 year

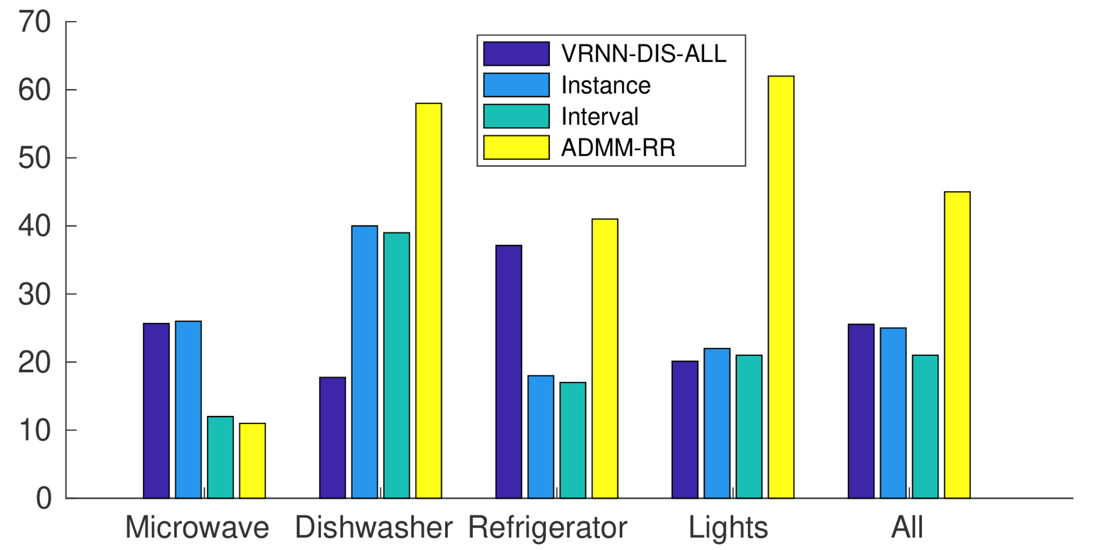
Preprocessing

1) Slide not overlapped window
2) Filtering moments with no activation
3) Split in train-validation-test: 0.5-0.25-0.25

MAE

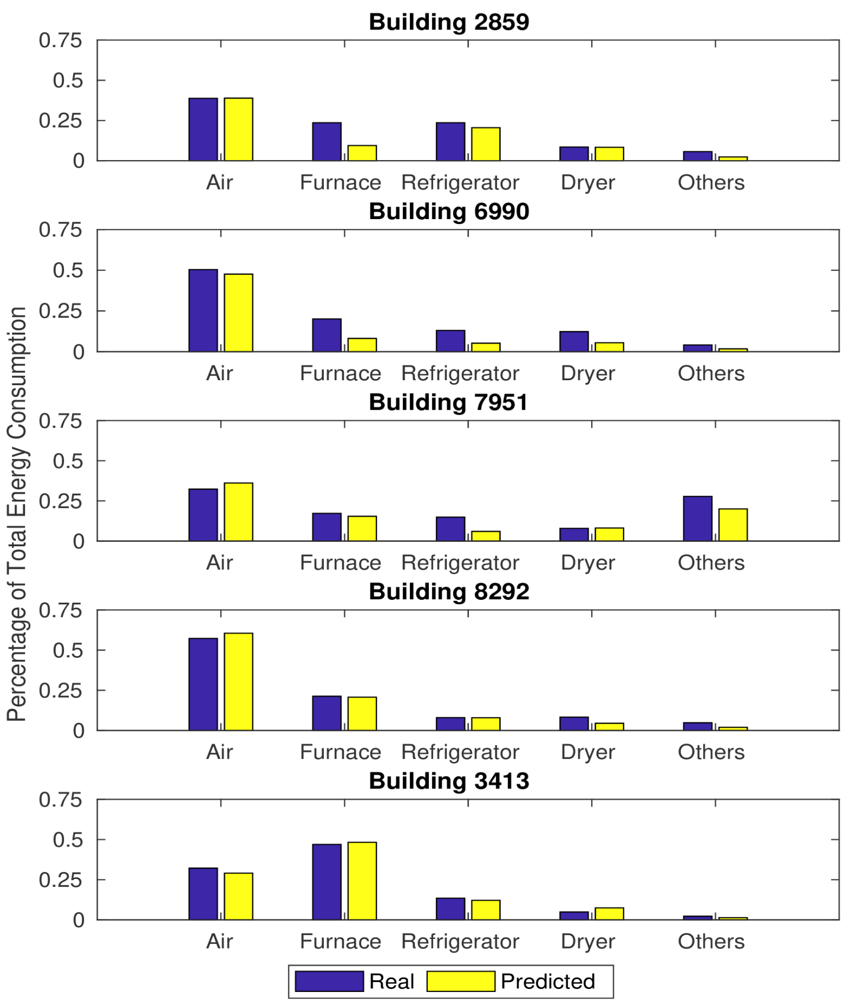
- Dataport (left): Our model outperforms all the other models in average across all the appliances and all the buildings
- REDD (right): Our model outperforms other for appliances such as dishwasher and lights



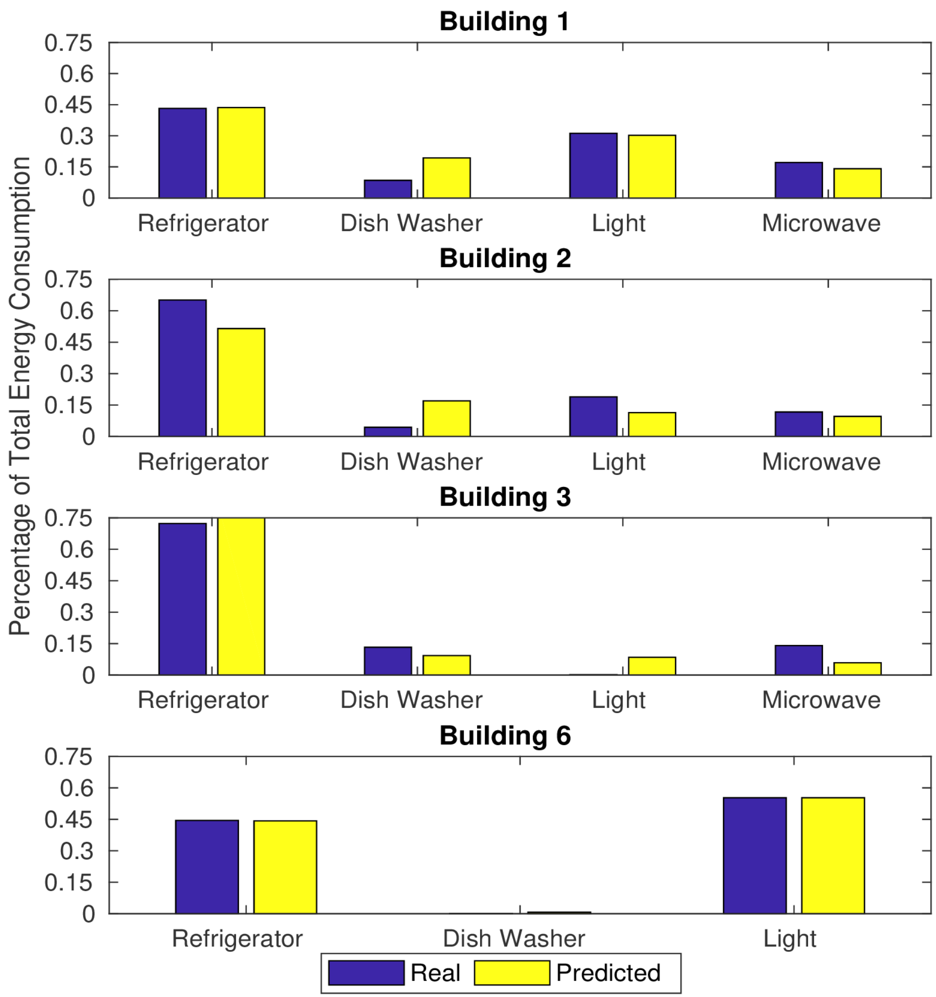


$$\%TotalEnergy = \frac{y^i_{1:T}}{x_{1:T}}$$

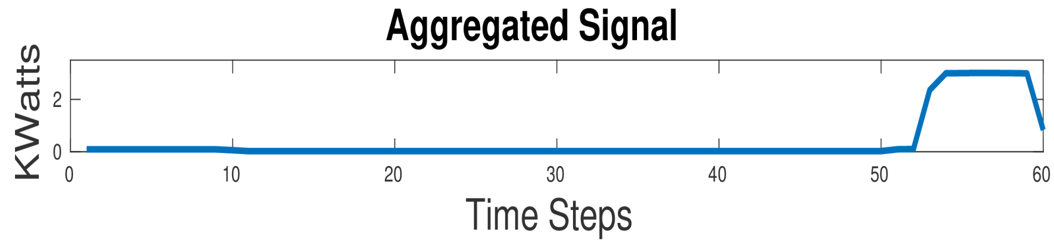
Dataport



REDD

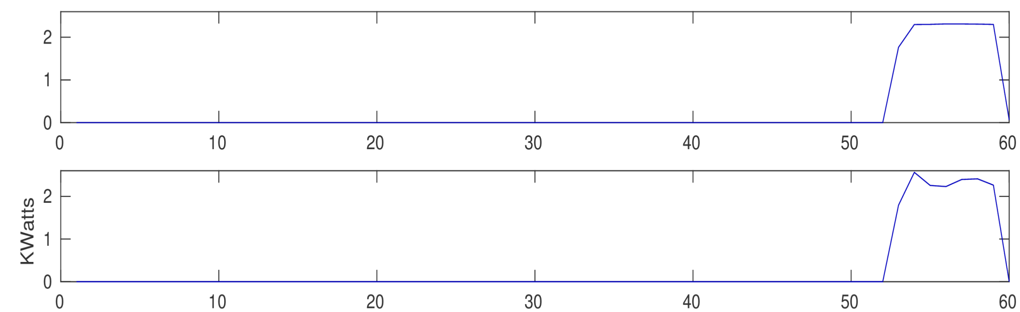


Dataport

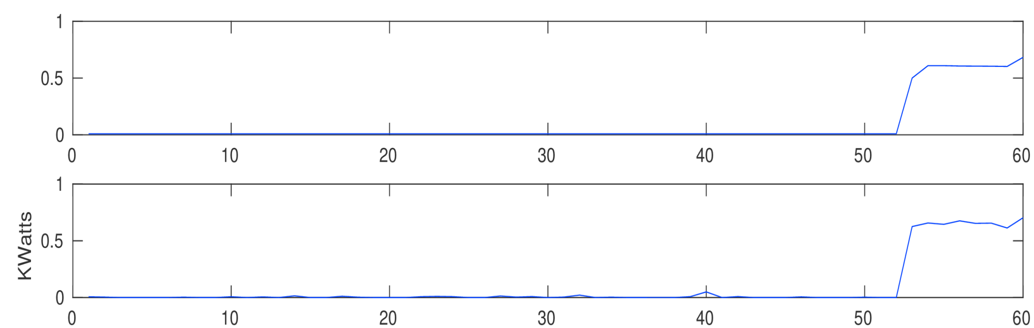


- Building 8292

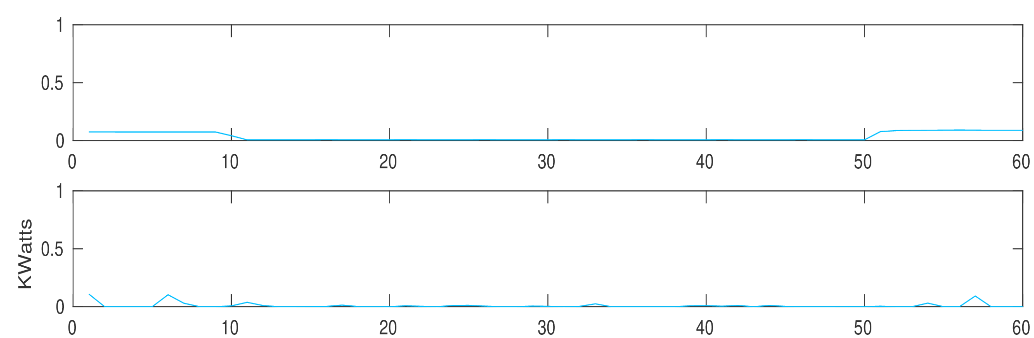
  - Air-conditioner



  - Refrigerator

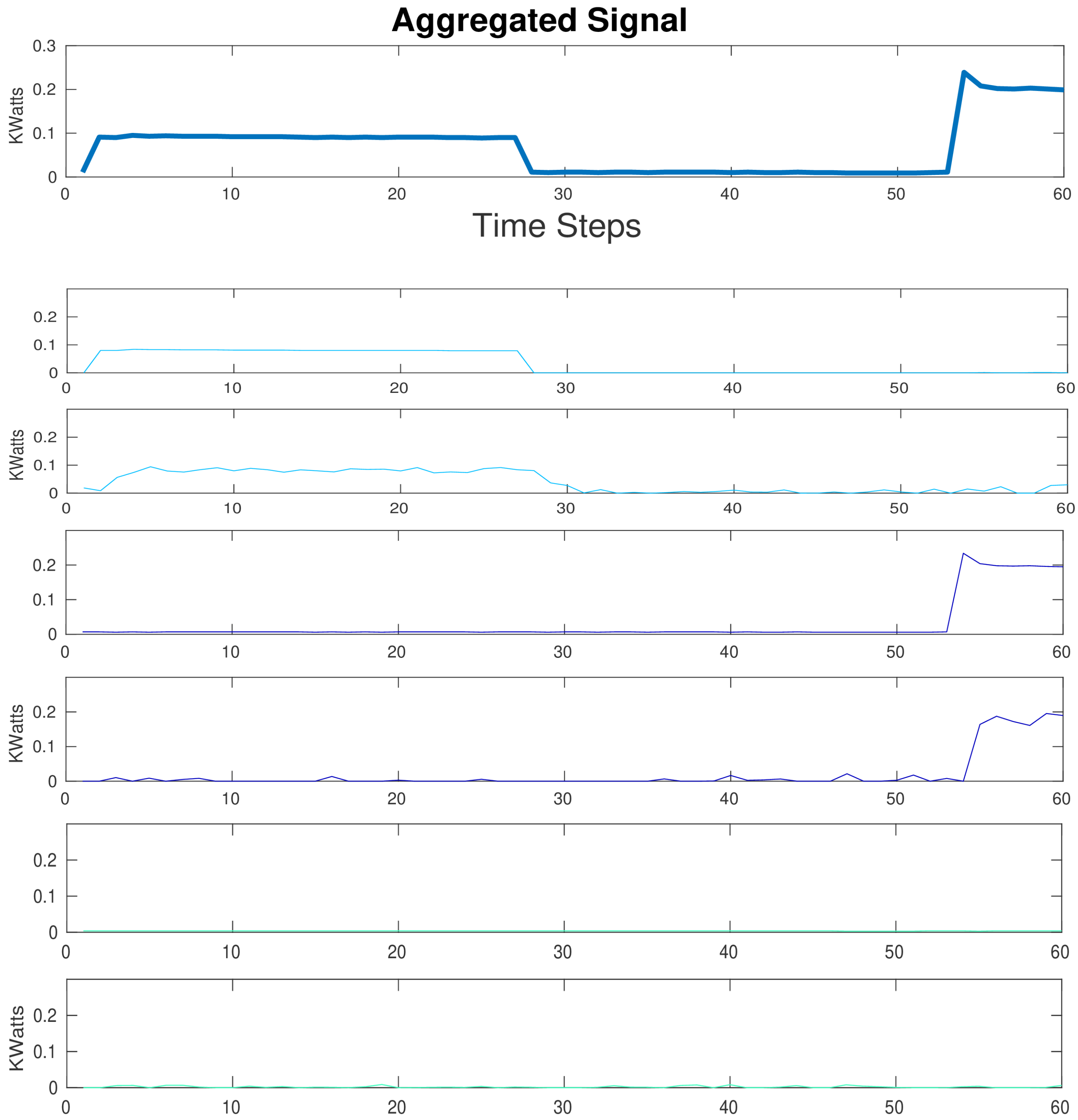


  - Furnace



REDD

- Building 1
  - Light
  - Refrigerator
  - Microwave



Thank you!